

I SAGGI DI LEXIA

48

Direttori

Ugo VOLLI

Università degli Studi di Torino

Guido FERRARO

Università degli Studi di Torino

Massimo LEONE

Università degli Studi di Torino

Aprire una collana di libri specializzata in una disciplina che si vuole scientifica, soprattutto se essa appartiene a quella zona intermedia della nostra enciclopedia dei saperi — non radicata in teoremi o esperimenti, ma neppure costruita per opinioni soggettive — che sono le scienze umane, è un gesto ambizioso. Vi potrebbe corrispondere il debito di una definizione della disciplina, del suo oggetto, dei suoi metodi. Ciò in particolar modo per una disciplina come la nostra: essa infatti, fin dal suo nome (semiotica o semiologia) è stata intesa in modi assai diversi se non contrapposti nel secolo della sua esistenza moderna: più vicina alla linguistica o alla filosofia, alla critica culturale o alle diverse scienze sociali (sociologia, antropologia, psicologia). C'è chi, come Greimas sulla traccia di Hjelmslev, ha preteso di definirne in maniera rigorosa e perfino assiomatica (interdefinita) principi e concetti, seguendo requisiti riservati normalmente solo alle discipline logico-matematiche; chi, come in fondo lo stesso Saussure, ne ha intuito la vocazione alla ricerca empirica sulle leggi di funzionamento dei diversi fenomeni di comunicazione e significazione nella vita sociale; chi, come l'ultimo Eco sulla traccia di Peirce, l'ha pensata piuttosto come una ricerca filosofica sul senso e le sue condizioni di possibilità; altri, da Barthes in poi, ne hanno valutato la possibilità di smascheramento dell'ideologia e delle strutture di potere. . . . Noi rifiutiamo un passo così ambizioso. Ci riferiremo piuttosto a un concetto espresso da Umberto Eco all'inizio del suo lavoro di ricerca: il "campo semiotico", cioè quel vastissimo ambito culturale, insieme di testi e discorsi, di attività interpretative e di pratiche codificate, di linguaggi e di generi, di fenomeni comunicativi e di effetti di senso, di tecniche espressive e inventari di contenuti, di messaggi, riscritture e deformazioni che insieme costituiscono il mondo sensato (e dunque sempre sociale anche quando è naturale) in cui viviamo, o per dirla nei termini di Lotman, la nostra semiosfera. La semiotica costituisce il tentativo paradossale (perché autoriferito) e sempre parziale, di ritrovare l'ordine (o gli ordini) che rendono leggibile, sensato, facile, quasi "naturale" per chi ci vive dentro, questo coacervo di azioni e oggetti. Di fatto, quando conversiamo, leggiamo un libro, agiamo politicamente, ci divertiamo a uno spettacolo, noi siamo perfettamente in grado non solo di decodificare quel che accade, ma anche di connetterlo a valori, significati, gusti, altre forme espressive. Insomma siamo competenti e siamo anche capaci di confrontare la nostra competenza con quella altrui, interagendo in modo opportuno. È questa competenza condivisa o confrontabile l'oggetto della semiotica.

I suoi metodi sono di fatto diversi, certamente non riducibili oggi a una sterile assiomatica, ma in parte anche sviluppati grazie ai tentativi di formalizzazione dell'École de Paris. Essi funzionano un po' secondo la metafora wittgensteiniana della cassetta degli attrezzi: è bene che ci siano cacciavite, martello, forbici ecc.: sta alla competenza pragmatica del ricercatore selezionare caso per caso lo strumento opportuno per l'operazione da compiere.

Questa collana presenterà soprattutto ricerche empiriche, analisi di casi, lascerà volentieri spazio al nuovo, sia nelle persone degli autori che degli argomenti di studio. Questo è sempre una condizione dello sviluppo scientifico, che ha come prerequisito il cambiamento e il rinnovamento. Lo è a maggior ragione per una collana legata al mondo universitario, irrigidito da troppo tempo nel nostro Paese da un blocco sostanziale che non dà luogo ai giovani di emergere e di prendere il posto che meritano.

Ugo Volli

SEMIOTICA E INTELLIGENZA ARTIFICIALE

a cura di

ANTONIO SANTANGELO, MASSIMO LEONE

contributi di

**FEDERICO BELLENTANI, ANNA DALL'ACQUA, GUIDO FERRARO
TULIO FERREIRA LEITE DA SILVA, FRANCESCO GALOFARO
GIANMARCO THIERRY GIULIANA, REMO GRAMIGNA, MASSIMO LEONE
ANTONIO SANTANGELO, ELSA SORO, SIMONA STANO
ZENO TOFFANO, UGO VOLLI, CRISTINA VOTO**



aracne



©

ISBN
979-12-218-0429-4

PRIMA EDIZIONE
ROMA 10 GENNAIO 2023

INDICE

- 9 *Introduzione*
di MASSIMO LEONE
- 19 *Riassunto articoli*
di ANTONIO SANTANGELO

IA E SEMIOTICA GENERALE

- 29 *I compiti principali di una semiotica dell'intelligenza artificiale*
di MASSIMO LEONE
- 45 *Some remarks on fakes and deepfakes. A new epistemic challenge*
di REMO GRAMIGNA
- 65 *Mito e realtà dell'IA: uno sguardo semiotico*
di GUIDO FERRARO
- 91 *Il mito delle due intelligenze e la semiotica*
di UGO VOLLI

IA E SEMIOTICHE APPLICATE

- 107 *Verso una semantica quantistica? Applicazioni al discorso religioso*
di FRANCESCO GALOFARO, ZENO TOFFANO

- 131 *Using transformer networks and tensive semiotics to improve sentiment analysis accuracy in tourism digital platforms*
di TULLIO FERREIRA LEITE DA SILVA
- 149 *How to build a chatbot. A semiotic and linguistic approach*
di ANNA DALL'ACQUA, FEDERICO BELLENTANI
- 171 *L'intelligenza delle I.A. come effetto di senso semio-narrativo nei giochi digitali. Una rivoluzione semiosofica*
di GIANMARCO THIERRY GIULIANA

NARRARE L'IA

- 197 *Fit Different! An Exploration of Apple Fitness+ Artificial Affects Machine*
di ELSA SORO
- 219 *I corpi (im)possibili dell'Intelligenza Artificiale*
di SIMONA STANO
- 239 *Essere umani al tempo dell'intelligenza artificiale. Narrazioni a confronto*
di ANTONIO SANTANGELO
- 267 *Interpretare il cambiamento climatico attraverso l'intelligenza artificiale. Una prospettiva semiotica*
di CRISTINA VOTO

CONCLUSIONI

- 285 *Tra semiotica e intelligenza artificiale*
di ANTONIO SANTANGELO
- 293 *Autori*

INTRODUZIONE

MASSIMO LEONE*

ENGLISH TITLE: *Introduction*

ABSTRACT: This paper introduces the collection of essays by proposing a long-term perspective on the impact of the advent of artificial intelligence on the way human societies and cultures manage meaning. Like the emergence of language, and then that of writing, the increasingly exponential development of artificial intelligence changes the conditions of production, circulation, and preservation of meaning, and even its status. Artificial intelligence, this introduction argues, both fascinates and disquiets, makes one smile in its first vagaries and dismay at its possibilities. There are two areas in particular in which it makes one dream and tremble at the same time: that in which it now far surpasses human faculties, as in the game of chess, and that in which it imitates them in an increasingly indistinguishable manner. Then there are specific domains in which these two dimensions merge, as in the production of images or the imitation of human language. The ending of this introduction, then, holds a great surprise for its reader, inviting the perusal of the essays collected in the volume.

Keywords: Semiotics, Artificial Intelligence, Bots, Meaning, Signification, Imitation

A un certo punto, non si è parlato d'altro. L'intelligenza artificiale era ovunque. Era nelle conversazioni quotidiane, al bar o in un taxi, e allora subito si trasformava in oggetto di previsioni e cliché, su come ci sorveglierà, su come cambierà la nostra vita, sui posti di lavoro che eliminerà. Ma entrava anche nei discorsi accademici, quelli delle scienze umane e sociali, dove aver letto qualche articolo di divulgazione consentiva di presentarsi come esperti, e fare illazioni di nuovo sull'impatto,

* Università di Torino.

sui pericoli, sui dubbi, con quella smorfia critica che i filosofi riproducono ormai con facilità, oppure lanciandosi in entusiastici proclami sulla svolta rivoluzionaria che la nuova tecnologia avrebbe rappresentato nella ricerca. Prima timidamente, poi sempre più impetuosamente, un discorso sovraccitato sull'intelligenza artificiale si è fatto strada a livello planetario, saldandosi con quello anch'esso effervescente sul digitale, e incrociando l'insieme delle calamità che affliggono il mondo, le sue preoccupazioni, ma anche i suoi desideri, le sue ambizioni.

Se con un gioco della fantasia si riuscisse a immaginare il momento d'invenzione e diffusione delle prime tecniche di scrittura, di scrittura alfabetica per esempio, forse si potrebbe cogliere una fibrillazione analoga, sebbene probabilmente più lenta e con una geografia più frastagliata e irregolare: d'un tratto — gli umani dell'epoca si saranno detti — come dal nulla sorgono segni che significano qualcosa, che non sono come i segni naturali che gli umani già interpretano — i fulmini, le nuvole, le impronte degli animali — bensì segni artatamente creati dall'uomo, che non a tutti significano ma solo a coloro che li sanno decifrare, e che non da tutti sono creati bensì solo da coloro che ne conoscono il codice, e lo utilizzano per gestire la memoria e la comunicazione, ovvero il tempo e lo spazio dell'informazione, quello che si tramanda dal presente al futuro e quello che si trasmette dal qui all'altrove.

Anche la scrittura dunque forse generò stupore, un senso di mistero e di timore reverenziale, e un proliferare di segni inarrestabile, che giunge sino a noi. Generò, e se ne hanno tracce nella storia della scrittura che riflette su sé stessa, inclusioni ed esclusioni, avversioni e desideri, e poi un modo nuovo di gestire la costruzione della lingua e della cultura attraverso il linguaggio. Gli studi di Ong su oralità e scrittura hanno fatto il punto su molte delle conseguenze di questo passaggio, e forse oggi bisognerebbe fare altrettanto con le nuove scritture che affiorano all'alba dell'orizzonte tecnologico contemporaneo.

In effetti non di semplici scritti si tratta, come avvenne con l'invenzione del romanzo, o di nuove tecniche, come fu il caso con l'avvento della stampa a caratteri mobili, bensì di un nuovo tipo di scrittura nel senso più ampio del termine, ovvero nel senso di una nuova testualità. L'intelligenza artificiale sconvolge perché non si limita a creare nuovo senso, ma cambia le regole del senso umano, così come nel passato

fecero prima il linguaggio come prodotto dell'evoluzione biologica e poi la scrittura come suo prolungamento nell'evoluzione culturale. Il linguaggio ha permesso alla specie umana di trasformare le articolazioni cognitive dell'intelligenza in strutture socioculturali condivise, che fossero non solo interne agli organismi ma anche esterne a essi, estrinsecate in segni che potessero dunque riprodursi e trasformarsi nel tempo e nello spazio. La scrittura, poi, ha consentito una stabilizzazione di questa matrice di articolazioni intersoggettive, costituendo la traccia di una memoria collettiva non più affidata alle memorie e alle intelligenze individuali, ma pur sempre in costante relazione con esse. Il linguaggio creò pensiero condiviso, la scrittura generò memoria collettiva, e l'intelligenza artificiale sta forse dando luogo a un'elaborazione comune del pensiero che, esattamente come avvenne per la memoria con la scrittura, si autonomizzi dai corpi. In questa cronologia immaginaria della storia della significazione umana, il linguaggio è pensiero che si rende autonomo dai corpi, la scrittura è memoria che se ne rende indipendente, mentre l'intelligenza artificiale consente all'elaborazione di pensiero e memoria di continuare non solo al di fuori dei corpi, e non soltanto lontano nello spazio e nel tempo da essi, ma anche indipendentemente da essi.

Con un'immagine metaforica, si potrebbe pensare al vero effetto dell'intelligenza artificiale come a quello di una mano che, staccandosi nottetempo dal corpo, come in un sogno, continui a scrivere indipendentemente dal corpo cui essa normalmente appartiene, e dalla mente che solitamente la guida, eppure facendolo in un modo che costantemente ricorda quel corpo, e ricorda quella mente. In ciò si origina l'aspetto al tempo stesso affascinante e inquietante di questa mano, che si attiva lontano da noi nel tempo e nello spazio, forse anche in una dimensione diversa rispetto a quella della veglia, in una specie di sogno — o di incubo? — computazionale, ma che simultaneamente ci imita, ci assomiglia, balbetta vagiti che ricordano il timbro della nostra voce.

L'iperattivazione nei confronti di questo nuovo modo di fare senso, poi, si deve anche al fatto che esso si esprime per geografie variabili, e soprattutto per diversi gradi di visibilità. Vi sono elementi d'intelligenza artificiale negli algoritmi che determinano i risultati delle nostre ricerche nella rete, per esempio, con effetti a volte molto dirompenti sui

nostri modi di vita e comportamenti quotidiani; l'intelligenza artificiale dietro l'apparire di volti e profili su una cascata di schermate di Tinder, per esempio, potrebbe aver condotto milioni di persone a cambiamenti spesso radicali nella propria esistenza, culminati a volte nella decisione di formare una coppia o una famiglia con un altro individuo incontrato per il tramite di queste ricerche. Si discute oggi ampiamente, e a volte persino stancamente, sulle storture che questi mezzani tecnologici possono introdurre, ma in fin dei conti si deve ammettere che forse le inclinazioni e i pregiudizi di una app di social dating non sono molto dissimili da quelli di un qualsiasi lenone; i sensali digitali stupiscono per quantità e rapidità, ma restano comunque nell'alveo delle tecnologie familiari, forse perché, in fin dei conti, l'ultima parola poi spetta sempre agli utilizzatori/alle utilizzatrici, o ai consumatori/alle consumatrici, i/le quali, sia pure influenzati/e e guidati/e da mille strategie algoritmiche, alla fine devono dare il loro impulso di libido alla punta del polpastrello che decreta il permanere o meno di un volto e di un corpo nel reame del desiderio.

Vi è, però, un'intelligenza artificiale che risulta assai più vistosa, che ci affascina tremendamente, e che al contempo — come si diceva — a volte sgomenta: in buona sostanza, si tratta di quell'intelligenza artificiale che o ci imita o ci sovrasta, oppure, nell'ipotesi più eclatante, fa entrambe le cose. L'emersione del linguaggio, poi della scrittura, il loro depositarsi in strutture non solo socioculturali ma anche materiali esterne e intersoggettive, hanno dato alla specie umana un sentimento di straordinaria elevazione e unicità; per millenni, abbiamo creduto di essere non solo una specie ma anche di essere speciali, differenti e al di sopra delle altre specie, unici e investiti del potere di dominare la natura; molte ideologie religiose sin dall'antichità hanno corroborato questo atteggiamento, il quale si è vieppiù acuito con la modernità, staccandosi dal suo orizzonte metafisico tradizionale e radicandosi, invece, nel sentimento di una supremazia tecnica: il mondo è nostro perché nessun'altra specie — o almeno così pare — riesce a trasformarlo come facciamo noi, a distruggerlo come facciamo noi.

A un certo punto, tuttavia, con un decorso iniziato già con i primi vagiti dell'informatica bellica, un calcolatore dotato d'intelligenza artificiale è riuscito a battere il campione del mondo di scacchi, ed

è divenuto invincibile. Gli scacchi sono oggi un dominio in cui, incontrovertibilmente, l'intelligenza artificiale sovrasta quella naturale. Tuttavia questo exploit ha meravigliato e impensierito solo in una certa misura. In primo luogo, perché negli scacchi si esprime una sfera limitata dell'intelligenza umana, quella scacchistica appunto, che non necessariamente è ciò che tutti universalmente ammirano e cui aspirano o si inchinano; essere un campione di scacchi è un merito ben visto socialmente, ma difficilmente si affiderebbe a qualcuno il governo di un paese solo perché muove assai bene i suoi pezzi sulla scacchiera, in quanto è parte del senso comune che l'intelligenza umana si esprima e si faccia valere in forme molto più complesse di quelle necessarie per vincere una partita di scacchi.

In secondo luogo, se all'intelligenza artificiale si è riconosciuta una superiorità in questo dominio, si è trattato comunque di farlo così come si riconosce a un cavallo di correre più velocemente di un essere umano, o a una mietitrice di operare più efficacemente di un individuo con falce in mano. In altri termini, in questo come in altri ambiti applicativi, la superiorità dell'intelligenza artificiale su quella umana è stata considerata come essenzialmente quantitativa, sebbene dai risultati straordinari, un fatto di rapidità e capacità di calcolo più che un gradino misterioso, tale da suscitare un misto di deferenza e sgomento.

Il panorama ha cominciato a produrre smottamenti quando questa stessa capacità e rapidità di calcolo hanno cominciato a essere applicate non all'ambito delle decisioni razionali in un sistema chiuso, quale sono, in definitiva, gli scacchi, bensì ad ambiti che coinvolgevano facoltà umane diverse, in contesti dalle variabili molto più numerose e ambigue. Un primo livello di stupore è stato determinato dalla riproduzione del riconoscimento, e in particolare del riconoscimento facciale. Il fatto che una macchina sia in grado di riconoscere un volto fra mille in una folla che si muove lungo la strada di una metropoli è certamente sorprendente, in quanto esalta una capacità essenziale per la vita sociale umana — quella d'identificare i volti altrui — a una dimensione cui nessun essere umano, e neanche i cosiddetti “super-riconoscitori” potrebbero attingere.

In definitiva, tuttavia, anche in questo caso restava molto margine per riaffermare — sempre più nervosamente, però — la peculiarità e in

fondo anche la superiorità dell'umano; da un lato, il riconoscimento facciale non faceva altro che potenziare quantitativamente la memoria e la vista umane, senza per questo migliorarle significativamente dal punto di vista qualitativo; dall'altro lato, sia pure ammettendo la superiorità quantitativa, rimaneva il fatto che il riconoscimento facciale era molto meccanico, nel senso che riconosceva un volto ma senza *conoscerlo*, e continuava a scontrarsi con molte difficoltà anche nella mera funzione del riconoscimento — per esempio, gli umani continuavano a essere molto più efficaci nel riconoscere volti in condizioni contestuali difficili, come scarsa visibilità o movimento del volto stesso.

Oggi il progresso tecnologico sta erodendo questo margine di sicurezza del primato umano, non solo perché l'intelligenza artificiale riconosce volti con una rapidità e con una sicurezza ineguagliate dagli umani — in contesti difficili, con la mascherina, con scarsa visibilità, in movimento, etc. — ma anche e soprattutto perché questa intelligenza artificiale di riconoscimento sta acquisendo una sempre più spiccata capacità umana di conoscenza, che passa dal riconoscimento del volto ma che dal volto trae anche moltissime altre informazioni, associandole a tante altre, e arrivando a conclusioni che, per quanto ipotetiche, vanno ben al di là dell'estensione e del grado di precisione di qualsiasi fisiognomica umana. In una misura che cresce in modo esponenziale, in effetti, il nostro comportamento nelle reti digitali lascia tracce e produce dati che, spesso con il nostro consenso, ma molto spesso anche ben al di là di esso, costruiscono un nostro gemello digitale che non solo ha il nostro stesso volto, riconoscibile da una macchina, ma si presta anche ad algoritmi che mettono in relazione le nostre scelte, le comparano nel tempo, le paragonano con quelle d'innomerevoli altri gemelli digitali, fino a giungere a conclusioni e previsioni che a volte ci fanno sorridere per la loro ingenuità, mentre a volte ci sorprendono.

Da un lato, infatti, il gemello artificiale che ci propone l'intelligenza artificiale applicata ai dati massivi soprattutto nelle grandi piattaforme di espressione e consumo assomiglia alle nonne dell'Italia meridionale, per cui se una sera andavi a cena da loro, e ti preparavano la parmigiana, e a te piaceva moltissimo, e ti sperticavi in lodi sul suo conto, in seguito poi ti toccava la parmigiana per i dieci anni a seguire, anche se magari ne avevi abbastanza, o se quella volta che l'avevi mangiata e ti

era piaciuta in realtà era solo perché avevi tanta fame, o se l'avevi elogiata solo per essere gentile, etc. Allo stesso modo, l'algoritmo che continua a proporci cucce per cani anche se non abbiamo un cane, e anche se in fondo avere un cane in casa ci fa orrore, solo perché una volta per sbaglio siamo incappati in un video di cuccioli, e vi ci siamo soffermati più del dovuto, ci fa un po' sorridere, e come sappiamo il sorriso serve spesso a stemperare le paure rispetto alla nostra integrità ontologica.

Molto più nervosa è infatti questa smorfia d'ilarità quando abbiamo l'impressione che "Facebook ci ascolta", o "Amazon ci ascolta", perché sembra fare su di noi e sui nostri gusti e desideri previsioni che sarebbero impossibili a un'intelligenza umana, salvo che non sia stata utilizzata, appunto, per origliare quello che segretamente confessavamo a un nostro amico circa i nostri desideri più nascosti. Invece, incrociando tracce digitali dei nostri comportamenti online con dati massivi attraverso intelligenza artificiale, adesso il mio computer mi conosce meglio di quanto possa farlo un parente prossimo, e mi propone regali molto più confacenti alla mia personalità.

Non c'è però dominio di sviluppo e applicazione dell'intelligenza artificiale che più affascini e insieme inquieti gli umani di quello dove emerge la capacità di abbinare la capacità di calcolo con quella d'imitazione. Fino all'avvento dell'intelligenza artificiale, le macchine erano più cognitivamente potenti dell'umano, ma in fondo incapaci d'imitarlo nei suoi tratti più distintivi; gli animali come il pappagallo o le scimmie riuscivano invece a imitare egregiamente alcuni tratti del comportamento umano, ma con una capacità cognitiva che pareva nettamente inferiore e comunque poco versatile; oggi l'intelligenza artificiale comincia a inquietare perché non solo sovrasta l'umano in potenza di calcolo ma comincia anche a imitarlo a perfezione, scardinando in forme sempre più spettacolari quella consapevolezza — o forse quella presunzione — di unicità e superiorità che caratterizza la specie e ne guida l'operato nei confronti della natura. Vi sono tre ambiti, in particolare, dove questa fusione di primato cognitivo e capacità d'imitazione sta producendo risultati insieme sorprendenti e inquietanti.

Il primo è quello della produzione d'immagini. L'intelligenza artificiale riconosce immagini della realtà, la conosce attraverso incrocio con dati massivi, ma comincia anche a produrne simulacri sempre più

indistinguibili dall'originale, dapprima in immagini statiche bidimensionali, poi in immagini in movimento, adesso sempre più anche attraverso artefatti tridimensionali, e addirittura — nella robotica per esempio — in artefatti tridimensionali in movimento. Ancora sorridiamo quanto c'imbattiamo in un *deepfake* di un attore famoso, in un simulacro prodotto da intelligenza artificiale, eppure forse sorrideremmo di meno se la stessa tecnologia venisse utilizzata per “scrivere” doppioni di noi che si aggirino per il mondo digitale indipendenti e indisturbati, come cloni ribelli o gemelli dispettosi.

Il secondo dominio è quello del linguaggio; di nuovo, ci conferisce un guizzo di superiorità il vedere come un chatbot produca risposte bizzarre a domande normalissime, ma forse non ci rendiamo conto di quanto esponenzialmente rapido sia il miglioramento di questa tecnologia, e di come ben presto anche in questo ambito ci sarà difficile riconoscere il discorso prodotto da umani da quello prodotto da intelligenza artificiale, e in un numero crescente di contesti discorsivi. Noi docenti siamo forse così sicuri che le tesine che riceviamo dai nostri studenti e dalle nostre studentesse siano unicamente il prodotto delle loro menti umane? Una volta un noto esponente politico italiano mi chiese di rispondere ad alcuni suoi quesiti sull'intelligenza artificiale, e io non feci altro che girargli le risposte prodotte da un chatbot, senza che il politico in questione si accorgesse minimamente della sostituzione fino a quando non gliela rivelai, suscitandone lo stupore e l'inquietudine. In effetti anche in questo caso si sta producendo un passaggio analogo a quello fra riconoscimento facciale e *conoscenza* del volto: non si tratta più solo di un'intelligenza artificiale che riconosce costrutti verbali e li traduce in un'altra lingua — anche in questo caso con risultati sempre più spettacolari — bensì di un'IA che produce i suoi propri costrutti verbali in interazione con noi, spesso arrivando a farci credere che di fronte a noi vi sia un altro umano e non una macchina.

Il terzo ambito, che in un certo senso combina e fertilizza i primi due, è quello della creatività; sfruttando dati massivi sulla relazione fra immagini e testi verbali, per esempio, l'intelligenza artificiale odierna produce scenari visivi sorprendenti sulla base di input verbali; il risultato che ne deriva è ancora piuttosto stereotipato — dirà qualcuno — eppure dall'aspetto abbastanza creativo da competere con un grafico umano

contemporaneo di media inventività, forse non con Michelangelo, ma certamente con un neolaureato in grafica digitale dalla formazione e dall'indole creativa spesso assai più stereotipata di quella di una macchina. In effetti, vi sarebbe molto da riflettere su come a questa crescente esuberanza creativa della potenza di calcolo computazionale sembri oggi corrispondere una standardizzazione progressiva delle pratiche umane di senso, come se vi fosse una tendenza secondo cui, dopo aver delegato la cognizione, la memoria, e l'elaborazione, il passo successivo possa consistere nel delegare all'intelligenza artificiale i compiti che gli umani compiono con maggiore creatività.

Di fronte a questo scenario, in cui potenza di calcolo e capacità d'imitazione danno luogo a nuove forme di (pseudo?)creatività meccanica, non è difficile chiamare in causa, fra le scienze umane e sociali, la semiotica, affinché dica qualcosa di pertinente e specifico a proposito delle nuove vie del senso aperte dall'intelligenza artificiale. Dire qualcosa di specificamente semiotico sull'IA rientra in effetti perfettamente nelle corde di una disciplina che sin dalla sua fondazione si occupa di significazione, di senso, di emulazione, di simulacri, e anche d'innovazione e creatività.

La semiotica è lo studio dei segni e dei simboli e di come vengono utilizzati per comunicare un significato. Nel contesto dell'intelligenza artificiale (IA), la semiotica può essere utilizzata per analizzare i modi in cui i sistemi di IA comunicano e i significati che vengono trasmessi attraverso le loro interazioni con gli esseri umani e altri sistemi. Un aspetto della semiotica particolarmente rilevante per l'IA è il concetto di "significante", che si riferisce alla manifestazione fisica di un segno, come una parola o un'immagine. Nel caso dell'IA, il significante potrebbe essere un programma informatico o un modello di apprendimento automatico, mentre il significato è il concetto o il senso che il significante rappresenta. La semiotica può essere utilizzata anche per esaminare i contesti culturali e sociali in cui i sistemi di IA vengono sviluppati e utilizzati e come possono essere percepiti e interpretati da diversi gruppi di persone. Ciò può essere particolarmente importante per comprendere i potenziali pregiudizi e i limiti dei sistemi di IA e il modo in cui possono essere percepiti e utilizzati dai diversi utenti. Nel complesso, la semiotica può fornire un quadro utile per comprendere le complesse

interazioni tra i sistemi di IA e gli esseri umani e il modo in cui queste interazioni si formano e sono modellate dalle norme culturali e sociali.

A proposito, il paragrafo precedente è stato creato dal chatbot ChatGPT di OpenAI in risposta alla domanda “What can semiotics say about artificial intelligence?”; la risposta è stata tradotta in italiano da DeepL. Forse non il paragrafo più originale e brillante di questa introduzione, ma forse anche il più chiaro. Buona lettura.

RIASSUNTO ARTICOLI

ANTONIO SANTANGELO

Il libro è suddiviso in tre sezioni. La prima è sulle problematiche di natura generale che l'intelligenza artificiale pone alla teoria semiotica, ma anche, viceversa, sul modo in cui alcune nozioni semiotiche molto attestate possono essere utilizzate, per contribuire allo sviluppo della stessa intelligenza artificiale. La seconda sezione, invece, verte sul contributo che la semiotica può fornire nei vari ambiti di applicazione dell'IA, dall'*information retrieval* alla *sentiment analysis*, dalla progettazione dei *chatbot* a quella dei videogiochi. La terza sezione, infine, è sulle narrazioni che vengono utilizzate per parlare di intelligenza artificiale, determinando il senso che quest'ultima assume, oggi, nella nostra cultura.

Per quanto riguarda la prima sezione più teorica, Massimo Leone la apre indicando i compiti essenziali di una semiotica dell'intelligenza artificiale: studiare come essa simuli l'espressione dell'intelligenza umana; se e come possa dirsi creativa; quali sono i presupposti ideologici da cui muove chi la produce e chi la utilizza. Lungo queste tre direttrici, si dipanano le riflessioni di tutti gli articoli che seguono. Più nello specifico, poi, Leone si concentra sulle reti neurali avversariali (GAN), che sembrano prospettare l'ingresso delle nostre società nel "regno del falso assoluto". In particolare, quando esse producono volti di persone mai esistite, oppure quando riproducono in maniera realistica quelli di individui esistenti, pongono problemi interpretativi di difficile soluzione, che devono e possono essere affrontati con gli strumenti della

semiotica, già da Eco definita come una “teoria della menzogna” (Eco 1975: 17).

Con questi ultimi temi si confronta anche Remo Gramigna, che si occupa proprio della “semiotica dei *deepfake*”, prodotti audiovisivi realizzati per mezzo dell’IA, simulando azioni e discorsi che sembrano ma non sono realizzati da persone vere. Gramigna prima ricostruisce una storia della rappresentazione mediata del reale, mostrando come i mezzi di cui l’uomo si è servito a questo scopo si siano andati via via affinando per riprodurre il mondo in maniera verosimile ma “inaffidabile”, allontanandosi da un rapporto di referenza denotativa rispetto a ciò che mostrano. Per questo, egli sottolinea come sia sempre più necessario tenere separate le sfere del verosimile e del vero, riflettendo piuttosto su quelli che in semiotica si chiamano “effetti di realtà”. Dunque, passa in rassegna gli studi dei semiologi e dei semiotici che si sono occupati di questi problemi, mette in evidenza categorie e strumenti di indagine che possono tornare utili per affrontarli, e infine conclude ponendosi alcune importanti domande circa i risvolti etici e pratici di una società in cui l’intelligenza artificiale rende e renderà sempre più semplice, alla portata di tutti, produrre contenuti mediatici dallo statuto ambiguo su sé stessi e sugli altri.

Guido Ferraro, invece, riprende le prime suggestioni di Leone, a proposito del rapporto tra semiotica e IA, e si interroga sulla nozione di “intelligenza” che si può evincere osservando il modo di funzionare delle macchine. Egli sottolinea come la capacità di queste ultime di riconoscere le cose e di classificarle, magari collegandole ad altre, al fine di compiere delle azioni dotate di senso, sia tipica di ciò che noi riconosciamo all’intelligenza umana. In effetti, secondo Ferraro, queste operazioni sono di natura squisitamente semiotica, dato che consentono di assegnare un significato a ciò che si ha di fronte, in funzione delle pratiche di chi le compie. Dunque, capire come l’informatica sia in grado di raggiungere questi risultati, confrontando i suoi algoritmi con le conoscenze sviluppate dalla scienza dei segni, può essere utile per far crescere entrambe le discipline. Allo stesso tempo, però, appare evidente che il confine tra l’uomo e la macchina dotata di intelligenza artificiale stia, da un lato, nella capacità di essere creativi, quando si operano tali classificazioni e collegamenti, dall’altro nell’essere consapevoli del senso che

tutto ciò assume. Anche qui, invece di tirare una linea di demarcazione netta tra l'intelligenza umana e quella artificiale, Ferraro si addentra in una serie di riflessioni che tentano di far dialogare la semiotica e l'informatica. Infine, proprio sulla scia delle indicazioni di Leone, a proposito della necessità di interrogarsi sull'ideologia legata all'utilizzo di macchine che ci appaiono intelligenti, egli si domanda come possiamo fare ricorso a questo tipo di strumenti, per leggere la realtà che ci circonda e mettere eventualmente in discussione il nostro modo di vedere le cose.

La sezione più strettamente teorica del libro si conclude con l'articolo di Ugo Volli, che a sua volta si dimostra interessato a capire con che tipo di intelligenza abbiamo a che fare, quando parliamo di "intelligenza artificiale". Partendo da un'analisi dettagliata delle definizioni dizionariali del concetto di intelligenza umana, egli dimostra che essa è un processo o un'attività destinata all'adattamento da parte di individui necessariamente imperfetti. Questo, secondo lo stesso Volli, ha indotto filosofi e teologi di ogni epoca a cercare di stabilire meccanismi esatti di produzione di una intelligenza artificiale, capace di formalizzare automaticamente la semantica e di trarne verità inconfutabili. Le macchine di cui ci serviamo oggi, tuttavia, non sono la realizzazione di questi tentativi, per una serie di ragioni spiegate puntualmente da Volli nel suo articolo, ma più modestamente appaiono come l'effetto di senso della classificazione automatica di grandi insiemi di dati, basata su esempi umani, che consentono un feedback del software. Questo, comunque, non inficia il grande interesse che questi strumenti e il loro funzionamento assumono, sia per la semiotica, sia per le scienze cognitive e per le neuroscienze.

Gli articoli della seconda sezione del libro, pur se incentrati su alcuni ambiti di applicazione dell'intelligenza artificiale, non trascurano comunque di muovere da una solida base teorica, per mostrare come la semiotica possa essere utilizzata per migliorare il funzionamento degli strumenti informatici. In particolare, Francesco Galofaro e Zeno Toffano portano avanti in queste pagine una ricerca che conducono da quasi dieci anni, su un algoritmo di loro invenzione che esporta nel campo dell'*information retrieval* alcuni concetti e modelli della teoria dei quanti: l'*entanglement* e la correlazione dovuta all'interferenza quantistica. Essi esplorano, in una maniera chiara e accessibile, le

implicazioni linguistiche e semiotiche di una teoria semantica di tipo, per l'appunto, quantistico, ricostruendo le relazioni tra il loro lavoro e un ampio campo di studi che mutua strumenti statistici dalla teoria dei quanti e li applica alle scienze umane e sociali. In particolare, mettono alla prova il loro algoritmo su un caso di studio: il dibattito sviluppatosi su Twitter, attorno all'enciclica *Laudato si* di Papa Francesco. Galofaro e Toffano dimostrano così, in maniera elegante e convincente, che l'applicazione degli strumenti quantistici è utile per capire come la citazione di un testo su un social network, quando viene collegato a certi *hashtag* e a certe tematiche, modifichi il suo profilo semantico originale. Un fatto, questo, di cui il loro algoritmo è in grado di rendere conto, consentendo alla macchina che se ne serve di operare un'interpretazione "intelligente" come quella di un essere umano, almeno da questo punto di vista.

Un tentativo simile a quello di Galofaro e Toffano viene portato avanti anche da Tulio Ferreira Leite da Silva, nel suo articolo su come la semiotica tensiva di Zilberberg (2006, 2012, 2019) possa essere utilizzata per migliorare il funzionamento degli algoritmi basati sull'intelligenza artificiale, in particolare i *transformer neural networks*, che operano la *sentiment analysis* sui social network. Ferreira Leite da Silva mostra come la suddivisione tradizionale che viene fatta in questo ambito, tra opinioni positive, negative e neutre espresse dagli utenti, non sia sufficiente per rendere la complessità dei loro discorsi, che invece devono essere suddivisi in sei categorie: "*redoubling*", "*saturation*", "*attenuation*", "*minimization*", "*extinction*" "*re-establishment*". Lo studioso brasiliano mette alla prova i suoi strumenti teorici, anch'egli in maniera convincente e promettente, su un corpus di recensioni ricevute su Tripadvisor da un noto ristorante della sua nazione.

Anna Dall'Acqua e Federico Bellentan, invece, provano a servirsi della semiotica per migliorare il funzionamento dei *chatbot* che fanno ricorso ad algoritmi di IA per simulare una conversazione "naturale". Essi si concentrano su tre dimensioni della progettazione di questo genere di strumenti: la definizione degli obiettivi dell'utente e dei *chatbot*, che a volte possono divergere; il tratteggio di un "utente target" o "modello", del suo contesto culturale e pragmatico, e delle possibili interpretazioni che egli può fornire a ciò che gli viene detto dalla macchina;

le *personas* che possono essere utilizzate, per dotare i *chatbot* di un'identità riconoscibile con cui gli esseri umani possano aver voglia di interagire. In tutti questi ambiti, Dall'Acqua e Bellentani mostrano come la semiotica strutturalista riconducibile a Greimas, quella interpretativa di Eco e la pragmatica possano tornare utili, per confezionare strumenti informatici capaci di proporre interazioni piacevoli e, soprattutto, sensate.

Gianmarco Thierry Giuliana si concentra invece sui videogiochi, dove l'intelligenza artificiale viene notoriamente utilizzata per simulare quella umana, in modo che i giocatori si possano divertire come se interagissero con i propri simili. Giuliana sostiene che l'impressione di confrontarsi con uomini o donne sia solo un effetto di senso e lo dimostra sia descrivendo il funzionamento delle varie IA che vengono utilizzate in questo ambito, sia riflettendo sullo statuto semiotico della nozione di intelligenza di uso comune, che si può evincere dai dizionari. In pratica, dato che da un'entità "intelligente" ci aspettiamo che sappia interpretare l'ambiente in cui si trova in una maniera tale da risolvere i problemi che esso le pone per raggiungere i propri scopi, i programmatori dei videogiochi, che definiscono a priori gli obiettivi che possono e devono essere perseguiti all'interno dei loro mondi, fanno in modo di creare algoritmi che si oppongono a tutto questo o che lo favoriscano in maniera sensata, e tanto basta per far apparire i personaggi sintetici da essi animati come "intelligenti". Nel suo articolo, Giuliana mostra quali sono i tipi di "intelligenza" macchinica con cui un giocatore si può confrontare nei videogames e arriva a sostenere che le IA sviluppate in questo ambito finiscono per apparire indistinguibili dagli esseri umani, per una serie di ragioni ben argomentate, che hanno a che vedere con il funzionamento specifico dei contesti videoludici.

La terza sezione del libro, sulle narrazioni dell'intelligenza artificiale, si apre con l'articolo di Elsa Soro, che riflette su come viene raccontata, oggi, dagli autori delle applicazioni digitali per il fitness, l'esperienza dello sport e degli allenamenti operati per mezzo di strumenti informatici "intelligenti". In particolare, dopo un'analisi dei servizi e della comunicazione dei competitor, Soro si concentra sulla app di Apple denominata Fitness+ e ne analizza funzionalità, interfaccia, modalità di utilizzo e strategie di marketing, per dimostrare come si cominci a

intravedere, da parte di un colosso come l'azienda di Cupertino, una strategia discorsiva finalizzata a non parlare dell'IA come del principale elemento innovativo che dovrebbe indurre le persone a servirsi dei suoi strumenti. In effetti, in questo ambito, molte altre applicazioni ricorrono all'intelligenza artificiale per analizzare i dati prodotti dagli utenti prima, durante e dopo le loro performance, in modo da migliorarle. Ma ormai è scontato pensare che la salute del proprio corpo possa essere misurata e presa in carico dalle macchine, dunque, per scegliere a quale IA affidarsi, si utilizzano altri criteri, come il tipo di socialità — politically correct e molto americana — a cui essa può dare accesso. Questo, infatti, sembra essere il vero *plus* che Fitness+ di Apple propone ai propri utenti.

Comunque, il tema della problematicità del rapporto tra l'intelligenza artificiale e il nostro corpo è lungi dall'essere superato dalle strategie di comunicazione di aziende come Apple. A questo proposito, Simona Stano si addentra nel mondo della medicina e in quello della robotica, ponendosi alcuni interessanti interrogativi: ha senso inserire l'IA nei nostri corpi per curarli o potenziarli? Quando lo facciamo, la consideriamo parte di noi? Riteniamo ancora di essere completamente "umani"? E quando invece dotiamo l'intelligenza artificiale di un corpo proprio, magari umanoide, questo ci dà l'impressione di relazionarci con qualcosa di più simile a noi, rispetto a una semplice macchina? La risposta è complessa e Stano la ricerca nella semiotica della corporeità, poiché, ancora una volta, come in molte altre parti di questo libro, la sua ipotesi di fondo è che tutto si basi su effetti di senso, determinati da certi modi di pensare l'uomo, i suoi artefatti e il rapporto che egli deve instaurare con questi ultimi.

Come è evidente, questi ragionamenti ci inducono a domandarci se le tecnologie che utilizziamo nella vita di tutti i giorni e, in particolare, quelle basate sull'intelligenza artificiale, stiano modificando l'immagine che abbiamo di noi stessi, l'idea della nostra "umanità". Del resto, molti saggi sul digitale e sulla IA si pongono fin dal loro titolo il quesito su ciò che significhi essere umani nel mondo ipertecnologico che stiamo costruendo (Harari 2015; Tegmark 2017; Fry 2018; Frischmann e Selinger 2019; Ferraris 2021). Antonio Santangelo, quindi, analizza questo genere di discorsi, mostrando quali sono i modelli di pensiero

da cui essi derivano. Si tratta di vere e proprie narrazioni — Santangelo ne individua di otto tipi, a cui si rifanno i saggi scientifici e le opere di finzione del suo corpus d'analisi — che riconoscono a uomini e macchine alcune caratteristiche specifiche, ruoli e valori che danno origine a diverse visioni del mondo. Fondamentalmente, esse si differenziano in base al fatto che, in termini greimasiani, l'intelligenza artificiale funge da attante Soggetto, Oggetto di valore, Anti-Soggetto, Destinante, Destinatario, Anti-Destinante, Aiutante o Opponente, rispetto agli esseri umani. Tali narrazioni si distinguono anche, per l'appunto, in funzione dei valori in gioco al loro interno, che nella maggior parte dei casi hanno a che vedere con il tema della libertà, sia degli uomini che, in senso più o meno lato, della IA. Il quadro che emerge è quello di un contesto culturale in cui è grande e preponderante la preoccupazione circa le capacità prometeiche dell'umanità di liberarsi dalle varie catene che la tengono avvinta, ma in cui qualcuno — a dire il vero, soprattutto i ricchi protagonisti della rivoluzione digitale in atto e i governi che ne vogliono trarre profitto — mantiene viva la speranza che proprio l'intelligenza artificiale rappresenterà metaforicamente quel fuoco mitologico che consentirà alla nostra specie di elevare la propria condizione e di evolversi, magari anche verso qualcosa di super-, trans- o oltre-umano.

Questo genere di riflessioni comincia a gettare luce sul tema dell'ideologia legata al modo in cui interpretiamo il ruolo dell'intelligenza artificiale nella nostra vita e nelle nostre società. In questa direzione muove anche Cristina Voto, che analizza il dibattito sui dati che vengono utilizzati dagli algoritmi di intelligenza artificiale per trarre conclusioni sul cambiamento climatico. In questo ambito, risulta particolarmente evidente un fatto risaputo tra gli scienziati, ma forse non nell'opinione pubblica, vale a dire che le misurazioni che si compiono per dimostrare la validità di una teoria circa il funzionamento "oggettivo" della realtà sono guidate dalla teoria stessa, che impone quale parte di quest'ultima deve essere studiata. In un certo senso, quindi, la scienza è "soggettiva" e si fonda su un modo di vedere il mondo che appare il più sensato a chi se ne fa portatore. Questo può apparire strano a chi ritiene che, grazie ai dati e a un computer "intelligente", si possa dire come stanno effettivamente le cose, eliminando del tutto la soggettività dell'uomo

nell'interpretarle. Sarebbe bello, forse, poter affermare che, senza ombra di dubbio e senza possibilità di contro-argomentazioni, poiché ce lo dice una macchina, il nostro pianeta si sta surriscaldando per via dell'azione antropica, ma non è così, poiché chi critica questa affermazione sostiene che essa si fonda su dati parziali, raccolti apposta per sostenerla. Questo tipo di problema, naturalmente, impatta anche sull'intelligenza artificiale in generale, che molti vorrebbero utilizzare come un potente strumento di conoscenza della realtà, capace di giustificare con "obiettività" scelte politiche di grande rilevanza. A questo proposito, Cristina Voto propone di occuparsi dei modelli interpretativi del mondo che stanno alla base della costruzione dei database su cui lavora l'IA, poiché da essi deriva tutto ciò che quest'ultima ci consente di comprendere e affermare.

IA E SEMIOTICA GENERALE

I COMPITI PRINCIPALI DI UNA SEMIOTICA DELL'INTELLIGENZA ARTIFICIALE^[1]

MASSIMO LEONE*

ENGLISH TITLE: *The Main Tasks of a Semiotics of Artificial Intelligence*

ABSTRACT: The article indicates the essential tasks of a semiotics of artificial intelligence: studying the way it simulates the expression of intelligence; the way it produces content that is creatively endowed; the ideological assumptions of artificial intelligence within the culture that produces it. Artificial intelligence is, from a semiotic point of view, the predominant technology of fakery in the current era. On the strength of its studies on the false, semiotics can therefore also be applied to the analysis of the fake that, in increasingly sophisticated forms, is produced through artificial intelligence and through the deep learning of neural networks. The article focuses on the adversarial ones, trying to highlight their ideological assumptions and cultural developments, which seem to indicate the entry of human societies and cultures into the 'realm of the absolute fake'.

KEYWORDS: Generative Adversarial Networks; Fake; Simulation; Undecidability of the Digital Fake

1. Lo studio degli artefatti simulativi

La semiotica dovrebbe concentrarsi sullo studio degli sforzi per simulare i comportamenti intelligenti dell'uomo attraverso dispositivi non

(1) Questo saggio è il risultato di un progetto finanziato dal Consiglio Europeo della Ricerca (ERC) nell'ambito del programma di ricerca e innovazione Horizon 2020 dell'Unione Europea (Accordo di sovvenzione n. 819649–FACETS).

* Università di Torino; Università di Shanghai; Università di Cambridge; Fondazione Bruno Kessler, Trento.

organici e non umani. Questa simulazione può avvenire a livello di espressione, di contenuto o di entrambi. A livello di espressione, l'attenzione si concentra sulla riproduzione inorganica dei segni che gli esseri umani associano all'intelligenza. Si veda, ad esempio, il progetto "Gaze", finanziato dalla Disney⁽²⁾, un robot che simula le espressioni umane e, inoltre, le emula in un'interazione faccia a faccia (Fig. 1).

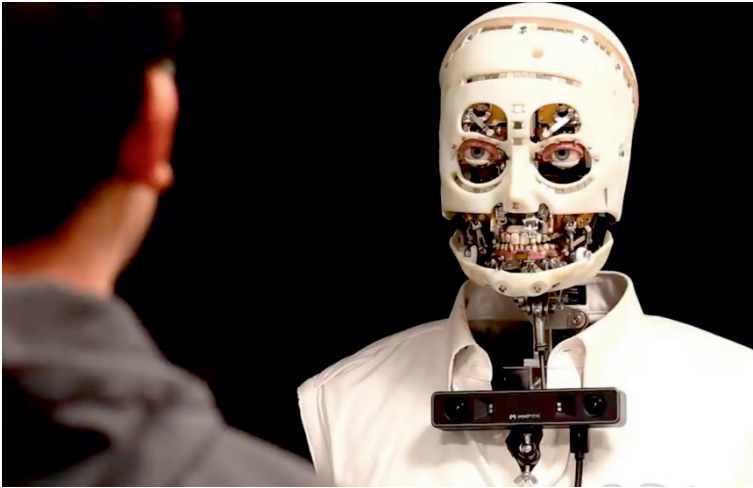


Figura 1. Il robot androide Gaze, finanziato dalla Disney, "interagisce" faccia a faccia con un essere umano; immagine di pubblico dominio.

In questa tecnologia, la materia essenzialmente inorganica viene organizzata, anche attraverso l'intelligenza artificiale, per trasmettere un'impressione di comprensione. Le espressioni facciali non sono solo cognitive, ma alcune lo sono e molte sono essenziali per comunicare la reciproca intelligibilità. A livello di espressione sono già stati fatti passi da gigante. Questo è evidente nel dominio del linguaggio verbale. L'espressione attraverso il linguaggio verbale è strettamente associata alle impressioni di intelligenza. Gli animali non umani tendono a non apparire dotati della stessa intelligenza e sensibilità degli esseri umani

(2) *Gaze*, un robot sviluppato da Walt Disney Imagineering con un gruppo di ricercatori dell'Università dell'Illinois a Urbana-Champaign e del California Institute of Technology, è il risultato di una ricerca avanzata nel campo della tecnologia e presenta una caratteristica molto interessante. È in grado di riprodurre alcune espressioni specifiche del volto umano. Ad esempio, è in grado di compiere piccoli movimenti della testa o di sbattere le palpebre.

perché non hanno accesso allo stesso linguaggio. Ma il linguaggio verbale può essere simulato a livello di espressione senza necessariamente comportare una simulazione a livello di contenuto. Per essere più precisi, l'intelligenza artificiale può simulare il linguaggio verbale attraverso una padronanza sintattica del linguaggio verbale umano. Possiamo parlare con il nostro altoparlante IA, come Siri o Alexa. Ma stiamo davvero parlando? Il nome stesso, "speaker", rivela che la tendenza ad attribuire comportamenti intelligenti umani ai dispositivi è antica. In passato, si trattava per lo più di una metafora. Un altoparlante tradizionale pre-IA non era in grado di parlare o di essere interpellato. Si limitava a riprodurre meccanicamente i suoni provenienti da una fonte. Ma un altoparlante di nuova generazione collegato all'intelligenza artificiale può, in una certa misura, interagire con gli esseri umani attraverso il linguaggio verbale. La padronanza sintattica del linguaggio verbale fornisce una solida infrastruttura per la simulazione dell'intelligenza. Tuttavia, è anche fondamentale che questa intelligenza ci parli attraverso una voce umana.

Una semiotica dell'intelligenza artificiale può applicarsi, in primo luogo, allo studio di tutti quei segni che vengono utilizzati per simulare l'intelligenza. A livello espressivo, i riferimenti al corpo umano sono fondamentali. Il progetto FACETS del Consiglio Europeo della Ricerca si concentra sul volto digitale per la sua attuale rilevanza sociale. Un aspetto di esso riguarda l'intelligenza artificiale. La simulazione del volto umano sta diventando sempre più centrale nella simulazione dell'intelligenza artificiale, soprattutto a livello espressivo. La simulazione di una pragmatica intelligente richiede la simulazione di un'estetica intelligente. Il fenomeno dell'*uncanny valley* ("la vallata inquietante") misura la resistenza all'estetica dei volti simulati, ma ne indica anche la potenza. In futuro, la simulazione di un volto intelligente sarà una caratteristica centrale dell'intelligenza artificiale. L'industria dei videogiochi, come sempre, è in anticipo. Un nuovo strumento di *MetaHuman* offre già agli utenti la possibilità di creare un'immagine digitale fotorealistica di un volto umano all'interno di un browser (Fig. 2). Epic Games ha annunciato questo software basato su browser e alimentato dal suo Unreal Engine⁽³⁾.

(3) L'azienda ha condiviso diversi video che illustrano questa tecnologia.



Figura 2. Un esempio d'immagine fotorealistica di volto creata dal browser di MetaHuman; immagine di pubblico dominio.

In un futuro più lontano, la generazione di volti digitali potrebbe combinarsi con quella di volti biologici. Come fanno notare Kim *et al.* (2020):

la rigenerazione di tessuti persi o danneggiati è l'obiettivo principale dell'ingegneria tissutale. Le tecnologie di bioprinting 3D sono state ampiamente applicate in molte aree di ricerca sulla rigenerazione dei tessuti e sulla modellazione delle malattie, con una risoluzione spaziale senza precedenti e una complessità simile a quella dei tessuti. Tuttavia, l'estrazione dell'architettura dei tessuti e la generazione di cianografie ad alta risoluzione sono compiti impegnativi per la rigenerazione dei tessuti. Tradizionalmente, tali informazioni spaziali sono ottenute da una raccolta di immagini microscopiche e poi combinate insieme per visualizzare le regioni di interesse. Per fabbricare questi tessuti ingegnerizzati, le immagini microscopiche renderizzate vengono trasformate in codice per informare un processo di bioprinting 3D. Se questo processo viene incrementato con approcci basati sui dati e ottimizzato con l'intelligenza artificiale, l'identificazione di un blueprint [formula] ottimale può diventare un compito realizzabile per la rigenerazione funzionale dei tessuti.⁽⁴⁾

(4) "Regenerating lost or damaged tissue is the primary goal of Tissue Engineering. 3D bioprinting technologies have been widely applied in many research areas of tissue regeneration and disease modeling with unprecedented spatial resolution and tissue-like complexity.

A un livello più profondo, l'intelligenza artificiale è sempre più applicata alla genomica e all'ingegneria genetica. I limiti alla ricreazione di specifici volti umani attraverso l'ingegneria genetica potenziata dall'intelligenza artificiale sono attualmente più etici che tecnici. I geni possono essere modificati, come nel caso dell'editing genetico CRISPR, una tecnica d'ingegneria genetica in biologia molecolare che consente di modificare i genomi degli organismi viventi⁽⁵⁾. Per il momento, l'intelligenza artificiale e l'apprendimento profondo stanno rapidamente progredendo nella simulazione di volti fotorealistici. La semiotica deve studiare la retorica dell'intelligenza artificiale, cioè i segni espressivi che simulano un comportamento o un dispositivo d'intelligenza (il volto è il dispositivo espressivo primario dell'intelligenza umana).

A livello di contenuto, l'attenzione si concentra sulla riproduzione inorganica del significato che gli esseri umani associano all'intelligenza. Il significato prodotto dall'intelligenza artificiale è, a un primo livello, comportamentista, cioè l'uomo si rende conto che le macchine sono intelligenti perché producono comportamenti intelligenti. Una calcolatrice può produrre nuovi risultati sulla base di input noti, ma difficilmente potrebbe essere considerata un dispositivo d'intelligenza. Il comportamento intelligente deve andare oltre il calcolo. I suoi risultati devono essere sintetici, non semplicemente analitici. Nei termini di C.S. Peirce, una macchina produce un comportamento che può essere qualificato come intelligente se coinvolge non solo la deduzione e l'induzione, ma anche l'abduzione. Anche in questo caso, tuttavia, la credibilità dell'intelligenza artificiale dipende dalla sua estetica. Una vecchia calcolatrice e un robot odierno sono entrambi dispositivi computazionali, ma il secondo è dotato di un'interfaccia simile a quella umana (una voce, un volto).

However, the extraction of tissue architecture and the generation of high-resolution blueprints are challenging tasks for tissue regeneration. Traditionally, such spatial information is obtained from a collection of microscopic images and then combined together to visualize regions of interest. To fabricate such engineered tissues, rendered microscopic images are transformed to code to inform a 3D bioprinting process. If this process is augmented with data-driven approaches and streamlined with machine intelligence, identification of an optimal blueprint can become an achievable task for functional tissue regeneration." (Trad. nostra).

(5) Si basa su una versione semplificata del sistema di difesa antivirale batterico CRISPR-Cas9. Introducendo in una cellula la nucleasi Cas9 complessa con un RNA guida sintetico (gRNA), il genoma della cellula può essere tagliato nel punto desiderato, consentendo di rimuovere i geni esistenti e/o di aggiungerne di nuovi in vivo (negli organismi viventi).

Più in generale, bisogna distinguere tra simulazione dell'intelligenza e intelligenza artificiale vera e propria. La prima è soprattutto una questione di segni espressivi, mentre la seconda richiede creatività a livello di contenuti. Ma l'intelligenza artificiale è creativa? A questo livello, quello dello studio dell'intelligenza artificiale come tecnologia produttrice di contenuti significativi, ciò che conta è definire l'intelligenza, i diversi tipi d'intelligenza e i modi in cui essi generano creatività. Esistono diversi modi in cui i contenuti creativi vengono creati dalle macchine attraverso l'intelligenza artificiale (nello specifico, il deep learning). Il più comune oggi è il riconoscimento di modelli nei big data. L'intelligenza artificiale è in grado di riconoscere configurazioni che inizialmente non vengono individuate dai ricercatori a causa delle dimensioni dei database su cui può operare. In passato, il vantaggio dell'intelligenza artificiale sull'intelligenza umana era in gran parte di tipo computazionale. I campioni di scacchi continuavano a vincere sulle macchine perché erano in grado di identificare e pianificare schemi di gioco che venivano ignorati dalle macchine. Ma ora la quantità sta diventando qualità.

Negli anni '80, all'apice della sua carriera, il campione del mondo di scacchi Garry Kasparov sosteneva che non sarebbe mai esistito un programma di scacchi in grado di sconfiggerlo. E in effetti, nel 1989, giocò due partite contro il computer *Deep Thought* dell'IBM, vincendole entrambe. Nel 1996, Kasparov sconfisse il successore di *Deep Thought*, *Deep Blue*, in una sfida su sei partite per 4 a 2, ma è stato il primo campione del mondo di scacchi a perdere una partita in condizioni di torneo contro un programma di scacchi. L'anno successivo, Kasparov fu sconfitto da *Deep Blue* nella rivincita. *Deep Blue* sorprese il mondo con un gioco "istintivo" e superiore che sembrava creativo sotto molti aspetti. Kasparov diffuse la voce che IBM doveva aver barato. Oggi, in 100 partite, il campione del mondo di scacchi in carica Magnus Carlsen non otterrebbe una sola vittoria contro il miglior programma di scacchi del mondo. Carlsen ha attualmente un rating Elo di 2.845 (febbraio 2019), mentre Stockfish 9 ha un rating di 3.438 (questi rating non sono FIDE, ma il parco giocatori computazionali è molto più competitivo di quello degli umani, per cui teoricamente un rating FIDE per Stockfish 9 dovrebbe essere ancora più alto). Le recenti prestazioni

dell'intelligenza artificiale nel tipico gioco cinese del Go sono ancora più spettacolari. Il gioco del Go è stato a lungo considerato il più impegnativo dei giochi classici per l'intelligenza artificiale, a causa del suo enorme spazio di ricerca e della difficoltà di valutare le posizioni e le mosse della scacchiera. Silver *et al.* (2016) introducono un nuovo approccio al Go al computer che utilizza “reti di valore” per valutare le posizioni sulla scacchiera e “reti di politica” per selezionare le mosse. Queste reti neurali profonde sono addestrate da un'inedita combinazione di apprendimento supervisionato da partite di esperti umani e apprendimento per rinforzo da partite di auto-gioco. Senza alcuna ricerca in anticipo, le reti neurali giocano a Go al livello dei più avanzati programmi di ricerca ad albero Monte Carlo che simulano migliaia di partite casuali di auto-gioco. Silver *et al.* (2016) introducono anche un nuovo algoritmo di ricerca che combina la simulazione Monte Carlo con reti di valori e politiche. Utilizzando questo algoritmo di ricerca, il programma AlphaGo ha raggiunto una percentuale di vittorie del 99,8% contro altri programmi di Go e ha sconfitto il campione europeo umano di Go per 5 partite a 0, essendo questa la prima volta che un programma informatico ha sconfitto un giocatore professionista umano nel gioco completo di Go, un'impresa che in precedenza si pensava fosse lontana almeno un decennio.

2. Lo studio delle ideologie simulative

Una filosofia della comunicazione digitale orientata alla semiotica mira a leggere le tecnologie del senso nel lungo periodo della storia dei sistemi semiotici umani, per rivelare le ideologie implicite che stanno alla base della creazione di nuovi dispositivi, processi e artefatti di significato. L'intelligenza artificiale non fa eccezione, poiché il suo sviluppo è solitamente sostenuto da preconcetti specifici su cosa sia l'intelligenza, su come dovrebbe funzionare e su quali tipi di risultati dovrebbe generare nel mondo. Poiché l'intelligenza artificiale sembra essere, almeno in questa fase del suo sviluppo, una simulazione dell'intelligenza umana, la semiotica può studiarla come un caso particolare di falsificazione, un tema che questa disciplina conosce bene. Ogni cultura e ogni

epoca storica sono caratterizzate dalle specifiche modalità semiotiche che adottano nella produzione del falso; l'intelligenza artificiale sta diventando la modalità principale nell'attuale produzione digitale del falso. La specie umana è dotata di una capacità innata di dare origine a rappresentazioni che intenzionalmente non corrispondono ad alcuna realtà ontologica. Le tecnologie e i linguaggi del falso, tuttavia, cambiano nel tempo e nello spazio. Con la tecnologia digitale, la comunicazione telematica e, soprattutto, con l'intelligenza artificiale e il deep learning, la cultura umana del falso sta varcando una soglia decisiva.

Nel mondo digitale, le culture umane entrano nel regno del "falso assoluto". Ciò è dovuto, in primo luogo, alle caratteristiche materiali della tecnologia digitale: tutto ciò che può essere oggetto di rappresentazioni digitali, può anche essere oggetto di rappresentazioni digitali senza riferimento ontologico. Qualsiasi immagine digitale che verrà prodotta del mio volto invecchiato in un futuro la cui ontologia non esiste ancora, può essere ricostruita nel presente da una simulazione digitale. In secondo luogo, il dominio del falso assoluto è causato dal potere dell'accumulazione quantitativa: un'immagine del mio volto ringiovanito può circolare sui social media in modo così intenso e virale che finirà per rappresentare la mia identità sul web. In terzo luogo, il dominio del falso assoluto è causato dalle sue nuove modalità di creazione: in precedenza, la falsità era una questione giocata tra contraffattori e intenditori (ad esempio, nel campo dell'arte); ora questa partita è giocata sempre più da algoritmi con risultati ampiamente imprevedibili. L'intelligenza artificiale applicata alla creazione del falso è sempre stata praticata nei confronti di un oggetto particolare, ovvero il volto, che è l'interfaccia principale e il dispositivo umano più importante per la comunicazione interpersonale.

La semiotica è perfettamente attrezzata per condurre uno studio il cui oggetto si trova all'incrocio tra il falso, il volto e le rappresentazioni digitali costruite dall'intelligenza artificiale. Per quanto riguarda il falso, tutti i padri fondatori della semiotica si sono occupati dell'argomento (Ousmanova 2004): 1) Charles S. Peirce nella tradizione americana (Cooke 2014); 2) le principali voci della semiotica strutturale, da un numero speciale della rivista francese *Communication* dedicato al concetto di "vraisemblable": Tzvetan Todorov, Gérard Genette,

Christian Metz, Julia Kristeva, Gérard Genot, Roland Barthes e altri (Todorov 1968); Baudrillard è tornato sull'argomento (1987; 2000); più recentemente, una tavola rotonda su "Post-verità e democrazia" è stata organizzata da Jacques Fontanille durante il Congresso dell'Associazione Francese di Semiotica a Lione, dall'11 al 14 giugno 2019 (Di Caterino 2020); Umberto Eco ha scritto molto sulla falsificazione (1995), ha curato un numero speciale della rivista semiotica *Versus* su "Fakes, Identity, and the Real Thing" (1987; con saggi di Eco, Prieto, Calabrese, *et al.*) e ha trattato l'argomento in numerosi saggi e romanzi (*Il pendolo di Foucault*, *Il cimitero di Praga*, *Numero Zero*); 3) Jurij M. Lotman ha più volte affrontato la questione del falso (Andrews 2003, p. 101; Makarychev e Yatsyk 2017). Anche il volto è stato oggetto di ricerca fin dagli inizi dell'antropologia moderna (Leone 2021). In semiotica, dopo i libri di Patrizia Magli sulla fisiognomica, il progetto ERC FACETS ha costantemente sviluppato la letteratura su questo tema.

3. Caso di studio: i volti dell'intelligenza artificiale

La ricerca semiotica sulle rappresentazioni digitali del volto sta crescendo sempre di più, soprattutto per quanto riguarda la rappresentazione del volto da parte dell'intelligenza artificiale. Per sviluppare un'analisi delle ideologie semiotiche che stanno alla base della creazione di volti sintetici, bisogna però guardare all'origine degli algoritmi che, negli ultimi anni, hanno rivoluzionato le pratiche in questo campo. Bisogna tornare al loro testo fondante, un articolo che il giovane Ian J. Goodfellow ha pubblicato il 10 giugno 2014 — quando era dottorando all'Università di Montreal — con il titolo *Generative Adversarial Nets*. Da allora, questo ricercatore è diventato un guru mondiale dell'intelligenza artificiale e in particolare del deep learning, ricoprendo posizioni di primo piano nel settore, tra cui quella di direttore del dipartimento di machine learning nel gruppo dei progetti speciali di Apple.

Insieme a un gruppo di amici dottorandi in informatica, Ian J. Goodfellow ha proposto un nuovo framework per la stima di modelli generativi attraverso un processo contraddittorio, in cui due modelli vengono addestrati simultaneamente: un modello generativo, che

cattura la distribuzione dei dati, e un modello discriminativo, che stima la probabilità che un campione provenga dai dati di addestramento piuttosto che dal modello generativo. Il modello generativo antagonista ha portato ad applicazioni rivoluzionarie nell'intelligenza artificiale e nel deep learning, tra cui la creazione di "volti artificiali" (Leone 2021) e di *deepfakes*. La semiotica è già stata applicata allo studio dell'intelligenza artificiale. Tuttavia, ha guardato ai suoi risultati e ai suoi prodotti, mentre sarebbe essenziale esaminarne, con la medesima prospettiva, i presupposti ideologici e la struttura soggiacente al suo funzionamento.

Lo schema di produzione dell'intelligenza artificiale immaginato da Goodfellow consiste in un'opposizione tra due istanze; il quadro della semiotica strutturale può quindi contribuire alla sua intelligibilità. Nell'architettura astratta delle GAN compaiono due attanti principali. Il primo è un attante generatore che esamina una configurazione di dati e produce un testo che potrebbe essere emesso da questa stessa configurazione; il secondo è un attante discriminatore che esamina il testo così prodotto e valuta se esso provenga dalla configurazione di dati o dall'attante generatore. Da un punto di vista epistemico, quindi, l'attante generatore mira a "far sembrare" e quindi a "far passare per vero" ciò che non lo è, mentre l'agente discriminatore mira a "far apparire" e quindi a "smascherare come falso" ciò che non è vero.

In termini matematici, per apprendere la distribuzione del generatore p_g sui dati x , si definisce un a priori sulle variabili di rumore in ingresso $p_z(z)$, quindi si rappresenta una mappatura nello spazio dei dati come $G(z; \theta_g)$, dove G è una funzione differenziale rappresentata da un *perceptron* multistrato con parametri θ_g . Viene definito anche un secondo *perceptron* multistrato $D(x; \theta_d)$, che produce un singolo scalare. $D(x)$ rappresenta la probabilità che x provenga dai dati piuttosto che dal p_g . D è formato per massimizzare la probabilità di assegnare l'etichetta corretta sia agli esempi di addestramento sia ai campioni di G . Contemporaneamente, G è formato per minimizzare $\log(1 - D(G(z)))$ (Fig. 3).

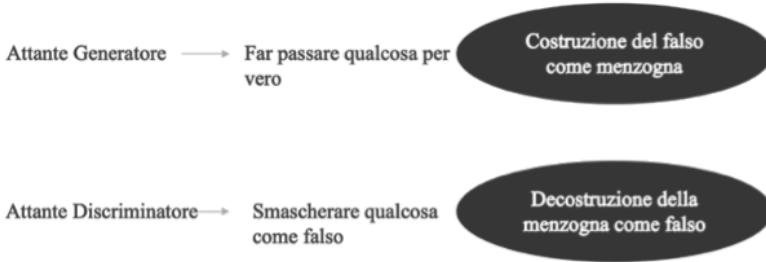


Figura 3. La dialettica attanziale nelle reti neurali avversariali.

L'attante discriminatore (D) è allo stesso tempo l'anti-soggetto dell'attante generatore (G), il suo destinatario e il suo coadiuvante. D sanziona i prodotti di G, designandoli come veri o falsi (cioè come provenienti o meno dall'insieme dei dati); quando la sanzione è positiva, però, determina ipso facto anche la sconfitta di D rispetto al suo anti-soggetto G, e viceversa: quando la sanzione è negativa, ciò porta alla sconfitta di G rispetto al suo anti-soggetto D. Ma in ogni caso D è sempre anche il coadiuvante di G, perché il secondo impara da ogni sanzione di D come ingannare meglio la sua controparte (Fig. 4).

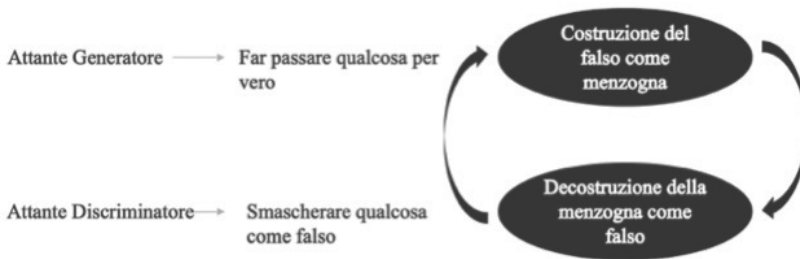


Figura 4. La spirale attanziale nelle reti neurali avversarie.

Quando si legge, attraverso la semiotica, l'articolo fondativo "Generative Adversarial Nets" (GAN), si rimane colpiti soprattutto da due elementi: 1) la concezione dell'intelligenza artificiale che esprime si basa sull'idea di antagonismo (né cooperazione, né semplice competizione); 2) la metafora che meglio spiega la nuova architettura del deep learning è quella del falsario e dell'intenditore (soprattutto nella

fabbricazione di denaro). Entrambi gli aspetti meritano un'ulteriore riflessione filosofica e semiotica, perché questa nuova architettura dell'intelligenza artificiale trova oggi applicazione in molti ambiti professionali e sociali, e in particolare nella creazione d'immagini e video sintetici di volti statici o in movimento, sempre più associati a teste, a corpi, oltre che a contesti sintetici, e che spesso comunicano attraverso molteplici sistemi di segni, come espressioni facciali, gesti, movimenti, frammenti di discorso verbale, canti, danze, ecc.

Il diagramma delle GAN può essere letto attraverso la metafora proposta dallo stesso Goodfellow nel 2014: D e G si comportano rispettivamente come un conoscitore e un falsario. Il falsario esamina la moneta in circolazione e cerca di produrne esemplari falsi; l'intenditore esamina la moneta prodotta dal falsario senza conoscerne l'origine e cerca di capire se si tratta di moneta falsa o autentica. Così facendo, però, il conoscitore fornisce al falsario informazioni che gli saranno utili per creare moneta falsa ancora più difficile da distinguere da quella autentica. Ma l'intenditore impara anche a discriminare sempre meglio tra denaro autentico e falso. Anche la metafora del mercato dell'arte può rendere l'idea di questa spirale di generazione e discriminazione: un falsario cerca di mettere in circolazione falsi Modigliani, mentre un conoscitore cerca di distinguerli dai Modigliani autentici; così facendo, però, il secondo fornisce al primo delle informazioni su come meglio falsificare le opere; viceversa, anche il primo impara dal secondo come falsificare le opere dell'artista italiano.

Ci si deve interrogare sulla natura dell'attante osservatore di questa spirale. I prodotti del modulo generatore non sono infatti sanciti solo dal modulo discriminatore, ma anche da un destinatario umano, che coincide, almeno in prima istanza, con il destinatario delle GAN. I moduli sono programmati da un mittente umano, ma i loro "comportamenti" non sono del tutto prevedibili, anche a causa del disallineamento computazionale tra cognizione umana e intelligenza artificiale. Il programmatore umano è quindi sia il mittente che il destinatario dei prodotti dell'interazione tra il modulo generatore e quello discriminatore. Inoltre, oltre a questo attore osservatore professionale, ce n'è un altro che è costituito da coloro che riceveranno i prodotti del modulo generatore senza conoscerne l'origine. La spirale appena descritta

finisce con l'aumentare sempre più l'incertezza epistemica di questo attore osservatore non professionista.

Per dirla in termini più semplici nel contesto della prima metafora: la competizione tra falsari e conoscitori mette in circolazione denaro o opere d'arte che sono falsi, ma che sono sempre più difficili da riconoscere come tali, soprattutto da parte dell'osservatore che si trova fuori dalla spirale. La circolazione massiccia di un falso non più identificabile come tale finisce per gettare discredito epistemico anche sulle opere d'arte e sulla moneta autentica. In questo risiede, forse, il pericolo più importante della "spirale del falso". Alcuni ricercatori hanno gettato una luce positiva sui GAN, suggerendo che la loro dialettica interna dovrebbe piuttosto essere paragonata a quella tra insegnante e studente. Il modulo generatore sarebbe quindi come uno studente che cerca di produrre rappresentazioni credibili da un insieme di dati, mentre il modulo discriminatore sarebbe come un insegnante che esamina e valuta queste rappresentazioni. Questo è in parte vero, ma ciò che fa la differenza è che, nel mondo delle GAN, le rappresentazioni del modulo generatore iniziano a circolare senza riferimento al contesto di apprendimento. Inoltre, è come se le GAN non consentissero all'allievo di superare definitivamente il maestro, obbligando il primo a condividere con il secondo il modo in cui ci è riuscito.

Questa è anche la differenza tra il falso digitale e quello analogico. La specie umana è intrinsecamente capace di produrre intenzionalmente false rappresentazioni della realtà, cioè rappresentazioni che, pur essendo prive di origine indicale, ne simulano una creando un effetto di senso iconico. Questa capacità è stata probabilmente selezionata dall'evoluzione biologica della specie come adattativa, in quanto ha permesso agli esseri umani di sperimentare mentalmente situazioni potenzialmente pericolose senza doverlo fare empiricamente. Inoltre, ha permesso loro di proteggersi dai predatori o di intrappolare le prede. Si tratta di una capacità che non manca in altre specie, sia vegetali che animali. Una delle peculiarità più notevoli degli uccelli liriformi, ad esempio, è la loro capacità di imitare i suoni, come quelli di altri uccelli e di vari elementi naturali, ma anche quelli dell'ambiente umano, come lo scatto di una macchina fotografica, il rumore di una motosega, il suono di un allarme antincendio, il frastuono di un cilindro idraulico, ecc.

Nella specie umana, tuttavia, questa capacità, espressa nel linguaggio e attraverso di esso, ha dato luogo a una sorta di “esattamento”, consistente nella capacità di attribuire piacere e valore estetico a rappresentazioni intenzionalmente false, che ha innescato a sua volta un’ enorme produzione di testi di finzione. La tecnologia digitale introduce un cambiamento qualitativo e quantitativo essenziale nella storia del rapporto della specie umana con il falso. Il digitale è dotato di una materialità proteiforme la cui manifestazione semiotica è completamente programmabile, cosa che non è mai avvenuta nella manifestazione dei testi predigitali. Ciò implica che qualsiasi rappresentazione digitale che abbia una relazione indicale con il suo oggetto può essere riprodotta in modo identico anche quando questa relazione è assente; la pittura può, ovviamente, simulare volti che non esistono, eppure lo scarto tra il volto ontologico e quello dipinto sarà sempre evidente, cosa che non accade nel digitale. Questo assorbe il senso di indicizzazione caratteristico della fotografia e lo riproduce in assenza di sostrato indicale; allo stesso tempo, introduce la piena programmabilità nella costruzione dell’immagine fotografica. Il quadro può rappresentare oggetti inesistenti ma non può far credere alla loro esistenza; la fotografia analogica può far credere all’esistenza degli oggetti che rappresenta, ma non può rappresentare oggetti inesistenti, almeno non efficacemente; la fotografia digitale può far credere all’esistenza degli oggetti inesistenti che essa rappresenta.

4. Conclusioni

L’applicazione dell’intelligenza artificiale, e in particolare del deep learning delle GAN, alla produzione della manifestazione materiale delle rappresentazioni digitali le sottrae alla valutazione umana. Il falso è inseparabile dalla specie umana, eppure è la prima volta nella storia della specie che agenti non umani sono stati messi nella condizione di produrre un falso la cui valutazione sfugge sempre più all’uomo ed è affidata sempre più a un esame che viene a sua volta effettuato per mezzo dell’intelligenza artificiale. I falsi digitali possono ormai riprodursi e circolare con una facilità senza precedenti, e questo aspetto quantitativo si traduce anche in un cambiamento qualitativo: è come se l’arte autentica

dovesse difendersi da un numero infinito di contraffattori che lavorano incessantemente e molto velocemente alla produzione di copie.

Il falso digitale è destinato, a lungo andare, a essere indistinguibile dal “reale digitale”; nel caso dei volti, ad esempio, è solo questione di tempo prima che non si possa più sapere dalla foto digitale di un volto se la foto sia stata prodotta da un volto biologico e ontologico o se si tratti di un’immagine sintetica. La semiotica tende a problematizzare il concetto logico di verità come adeguatezza al reale, considerando piuttosto le condizioni semiotiche che producono un “effetto di realtà”. Tuttavia, spiegare la retorica dell’effetto di realtà senza postulare una realtà ontologica conduce a ineludibili aporie. Allo stesso modo, si può ben problematizzare l’effetto di realtà di una fotografia analogica, ma si deve anche riconoscere che l’avvento del digitale, e del deep learning digitale applicato alla creazione di immagini, mina la possibilità di distinguere tra un’immagine referenziale dotata di un effetto di realtà e un’immagine sintetica che produce lo stesso effetto.

Riferimenti bibliografici

- ANDREWS E. (2003), *Conversations with Lotman: Cultural Semiotics in Language, Literature, and Cognition* [Toronto Studies in Semiotics and Communication], University of Toronto Press, Toronto, Buffalo e Londra.
- BAUDRILLARD J. (1987), *Au-delà du vrai et du faux, ou le malin génie de l’image*, “Cahiers internationaux de sociologie”, nouvelle série, numero speciale su “Nouvelles images, nouveau réel”), gennaio–giugno, 82: 139–45.
- (2000), *The Vital Illusion*, The Wellek Library Lectures, New York.
- COOKE E.F. (2014), “Peirce and the ‘Flood of False Notions’”, in T. THELLEFSEN, B. SØRENSEN, C. DE WAAL (a cura di) (2014), *Charles Sanders Peirce in His Own Words: 100 Years of Semiotics, Communication and Cognition* [Semiotica, comunicazione e cognizione 14], De Gruyter Mouton, Boston, 325–31.
- DI CATERINO A. (2020), *Fake News: Une mise au point sémiotique*, online, “Actes Sémiotiques”, 123; disponibile nel sito <https://www.unilim.fr/actes-semiotiques/6445> (ultimo accesso il 14 maggio 2022).
- ECO U. (1984), *On Fish and Buttons: Semiotics and Philosophy of Language*, “Semiotica”, 48, 1–2: 97–118.

- (1987), *Fakes, Identity and the Real Thing*, numero speciale di *Versus*, 46, Bompiani, Milano.
- (1995), *Faith in Fakes: Travels in Hyperreality* (1986), Minerva, Londra.
- (1975), *Trattato di semiotica generale*, Bompiani, Milano.
- KIM J. *et al.* (2020), *Engineering Tissue Fabrication with Machine Intelligence: Generating a Blueprint for Regeneration*, “Frontiers in Bioengineering and Biotechnology”, 7, 443: online; DOI: <https://doi.org/10.3389/fbioe.2019.00443>.
- LEONE M. (a cura di) (2021), *Volte artificiali/Artificial Faces*, numero speciale di *Lexia*, 37–38, Aracne, Roma.
- MAKARYCHEV A.S., YATSYK A. (2017), *Lotman’s Cultural Semiotics and the Political: Reframing the Boundaries*, Rowman & Littlefield International, Londra.
- OUSMANOVA A. (2004), *Fake at Stake: Semiotics and the Problem of Authenticity*, “Problemos”, 66, 1: 80–101.
- SILVER D., HUANG A., MADDISON C. *et al.* (2016) *Mastering the Game of Go with Deep Neural Networks and Tree Search*, “Nature” 529, 484–89; DOI: <https://doi.org/10.1038/nature16961>.
- TODOROV T. (a cura di) (1968), *Recherches sémiologiques le vraisemblable*, numero speciale di *Communications*, 11.

SOME REMARKS ON FAKES AND DEEPFAKES A NEW EPISTEMIC CHALLENGE⁽¹⁾

REMO GRAMIGNA*

TITOLO IN ITALIANO: Alcune considerazioni sui falsi e sui *deepfakes*. Una nuova sfida epistemica

ABSTRACT: I *deepfakes* sono video e/o audio che sono stati realizzati e manipolati attraverso tecniche di apprendimento automatico, in particolare reti generative avversarie (GAN). A causa dell'offuscamento della distinzione tra probatorio e finzionale, reale e virtuale, finto e falso, queste nuove forme di media sintetici funzionano sia come potenziale forma di disinformazione e inganno che come forma inedita di creatività. Sebbene si tratti di un fenomeno oggi abbastanza diffuso, la ricerca semiotica deve ancora mettersi al passo con gli studi in questo settore. I *deepfakes*, infatti, pongono sfide ontologiche, gnoseologiche ed epistemologiche che meritano un attento scrutinio.

KEYWORDS: deepfakes, artificial intelligence, fake, synthetic media, falsity

1. Introduction

Forgeries, fictions, illusions, and deception are very complex phenomena that have concerned humankind for centuries. Many renowned scholars dealt with the question, especially the philosophers — from St. Augustine to Immanuel Kant — who focused on why it is morally reprehensible to deceive others. Today this topic has regained momentum. It has become the subject of heated discussions, arousing interest

(1) This paper is the result of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement n. 819649–FACETS).

* Università degli Studi di Torino.

among experts — especially media scholars (Filimowicz 2022; Floridi 2018; Vaccari; Chadwick 2020; Schick 2020), computer scientists (Albahar, Almalki 2019; Karnouskos 2020; Tolosana et. Al 2016) and legal scholars (Meskys et al. 2020) — as well as the laypersons.

The “deepfake” in particular, a new technique able to create by means of deep learning methods very realistic videos that are almost indistinguishable from the authentic ones, has created quite a lot of debate in respect to the potential harm and the erosion of trust in news media that it may cause. This issue has divided the public opinion in two opposite fronts: the tech-enthusiast, who sees an unprecedented technology that may enhance education, arts and film industry (Kerner and Risse 2021)⁽²⁾, and the techno-phobic, who worries about the potentially negative outcomes⁽³⁾. The phenomenon of deepfakes opens up a vast array of issues with ramifications in many fields, from digital forensics to artificial intelligence. The present notes are not meant to be a fully-fledged account on deepfakes. These are rather some preliminary remarks on a subject that is new and thought-provoking. Undoubtedly, deepfakes would deserve a much lengthier and in-depth discussion.

2. An epidemic of images

In the last 30 months, there has been a huge shift in people’s life. For a host of different reasons related to the political and historical milieu we dwell at this current time, people have been plunged into digital communication more than ever. Both for work and education, conferences and social life as well as for entertainment and leisure, people spent more and more time online. Although we do not know what is going to happen in the near future, it is a plausible scenario that this trend may continue in the days ahead.

It is also worth pointing out that in the last two years, due to the Covid-19 pandemic, the use of online communication and social

(2) In this respect, see: <https://www.thinkautomation.com/bots-and-ai/yes-positive-deepfake-examples-exist/>

(3) See, for instance, the article published by the Washington Post, by Monkey Page, “Deepfakes are here. These deceptive videos erode trust in all news media”: <https://www.washingtonpost.com/politics/2020/05/28/deepfakes-are-here-these-deceptive-videos-erode-trust-all-news-media/>

media has skyrocketed as well as the sense of loneliness, with an increasing loss of meaning in people's life (Desmet 2022). Undoubtedly, in these years, the Sars–Covid–19 pandemic has been a driving force towards the consumption of online experiences. A reduction in face-to-face encounters is coupled with an increase in online interactions, including the use of virtual experiences, augmented reality and gaming. Much more time has certainly been spent at home, as in the last couple of years government regulations have drastically limited the individual movements and lifestyles of many citizens during prolonged lockdowns (Gramigna, Voto 2021; Matulewska, Wagner, Marusek 2022). On the other hand, remote working — the so-called “smart working” — has ushered in a trend that will probably be destined to last over time. Smart working is another piece that adds to the global push that, albeit dictated by necessity, has channelled many people towards a massive use of online activities and digital experiences. Social media and online environments have increasingly become the arena where people rely to acquire information and a means to shape the public opinion (Sunstein 2007; Pariser 2011).

If this is not enough to contend with, there has been a growing awareness around artificial intelligence (AI), a field that is growing at a fast pace. AI has had a huge impact on the creation, processing, and the circulation of images. In recent years, there has been a growth, a quantitative growth of digital images, some of which are synthetic images created through AI means and algorithms. This quantitative growth is unprecedented. Probably never in history the diffusion of images has reached such a pinnacle, so much so that we could conceive of it as an ‘avalanche’ of images. As Joan Fontcuberta convincingly argues:

For the first time we are both producers and consumers of images, and the simultaneous accumulation of these factors has resulted in an almost endless iconic avalanche. (...) We are thus passing through an era in which images are included in the category of excess, an era in which we speak of a mass production that leads to asphyxiation rather than emancipation. (...) The idea of the excess of images, which seems to produce at the same time hypervisibility and universal voyeurism, this apparent “epidemic of images”, deserves a more in-depth and critical treatment that involves not only the specific conditions of

the image, but also the logics of its management, dissemination and control.⁽⁴⁾

In a certain sense, deepfakes are a consequence of the massive overproliferation of images that we witness today. Without the massive amount of data available on the web, its relatively easy access, and the possibility of manipulation, it would be hard to imagine the birth of deepfakes content. Whilst in the past a unique object was the model for fakes and copies, usually few in numbers, today fakes are created *ex-nihilo* from pre-existing data sets, that is, already existing images. Deepfakes techniques create very realistic digital images, videos and audio material that are almost indistinguishable from the real ones.

There are a few things which I would like to address in this essay. To start with, one may wonder whether the time we live in marks a new era whereby fakes and forgeries are predominant as compared with earlier epochs. There is an important corollary to this question: what sets asides such new forms of digital fakery as compared to the previous and more traditional forms of image manipulation? Furthermore, on a more theoretical level, what has a semiotics of fakes achieved so far, what can we learn from it, and what are the future challenges ahead? Last but not least, deepfakes and digital manipulation of face images have pivotal ethical, social, epistemic, and semiotical ramifications that need to be spelled out.

3. Foreseeing the change

In semiotics, there have been a few scholars who foresaw that a radical shift in image representation was forthcoming. Rather than a systematic analysis, except for U. Eco's treatment of fakes (Eco 1990, pp.

(4) "Per la prima volta siamo produttori e consumatori d'immagini, e l'accumulo simultaneo di questi fattori ha provocato una valanga iconica quasi infinita. (...) Attraversiamo quindi un'era nella quale le immagini sono comprese nella categoria dell'eccesso, un'epoca in cui parliamo di una produzione di massa che porta come conseguenza l'asfissia piuttosto che l'emancipazione. (...) L'idea dell'eccesso di immagini, che sembra produrre nello stesso tempo ipervisibilità e voyeurismo universale, quest'apparente 'epidemia delle immagini', merita un trattamento più approfondito e critico che non implichi sono le condizioni specifiche dell'immagine, ma anche le logiche della sua gestione, diffusione e controllo" (Fontcuberta 2018, p. 53).

162–188), their accounts are more like a presentiment of what was about to come. Indeed, since the 1970s, the French–Lithuanian semiologist Algirdas J. Greimas (1989), pointed out that humanity entered in an era in which “symbolic manipulation” became the dominant paradigm. The corollary to this paradigm is that the boundaries between plausibility and veracity, reliability and certainty became blurred. In other words, the distinction between what is considered as “true” or “false” at a given historical moment and social context is not fixed, but constantly negotiated and questioned. This is not only a matter of cultural, contextual, and historical relativity, but a problem of steering and controlling what is regarded as plausible through discourse. From this perspective, foreseen by Greimas, truthfulness, plausibility, veracity are *effects of meaning*. These effects can be steered, controlled, and constructed through various textual and discursive strategies, which semiotic investigations must be able to uncover. Paraphrasing the famous article penned by Bolinger (1973), one could say that “truth is a semiotical question”.

When Greimas published his pivotal article, *The veridiction contract* (1989), he was writing in a time when the digital culture was not yet fully developed as we know it today. In the years since Greimas’ pioneering essay was written, technological awareness has grown at fast pace and breakthrough innovations were made. Indeed, today the phenomenon of symbol manipulation has been magnified and accelerated by the introduction of virtual reality, the widespread use of digital media, and the extraordinary growth in the field of AI.

Another, perhaps lesser-known author, who foresaw this epochal change is Jurgen Ruesch, an American psychiatrist and contributor to several important works together with Gregory Bateson. Ruesch referred to a phenomenon called “symbol manipulation”. He pointed out that through digital media, communication, messages and images have become detached from their original sources. A direct consequence of this state of affairs is that, although not impossible, it becomes increasingly difficult to assess the veracity of sources and to verify their reliability. I quote from a classic of semiotics, *Semiotic Approaches to Human Relation* (1972). It is an excerpt from a section entitled “Symbol Manipulation and the Representation of Reality”:

Through machine communication, messages have been detached from their originators and the truthfulness of the speaker no longer can be assessed. The erstwhile tightly controlled relationship between symbols and what they stood for has given away to a loose relationship whereby the number of possible interpretations has greatly increased. In a culture where messages are manipulated for advertising and propaganda purposes, the receiver is not quite sure of the sender's intentions and does not know whether he is dealing with a denotative or a connotative, a reality-oriented or a manipulative, a public or a personal statement. As a result, symbols are no longer trusted and therefore their informative value has been greatly reduced (Ruesch 1972, p. 268).

Thus, digital media have brought in radical changes in the use of symbolic systems and the representation of reality. This issue poses an important epistemological question, that is, how we acquire knowledge of reality and whether such a knowledge is reliable at all.

With Greimas and Ruesch we are still in the early days of such a semiotic consciousness. These authors foresee the change, but it will later be the post-modernists, like Jean Baudrillard, who will strongly thematise this aspect, speaking of the "society of simulacra". As early as 1995, Baudrillard foresaw the radical change introduced by digital images, which created what the author termed as *Integral Reality* or *Virtual Hyperreality*. According to Baudrillard (2004), the digital image — the simulacrum — sanctions the death, as it were, of the referent, its disappearance and nullification.

In a very lucid essay, Giovanni Gurisatti takes up Baudrillard's thesis and sets out the argument that not only is possible to conceive of the "perfect crime" (a crime without a trace), but we have reached a stage where the use of synthetic images yields to the possibility of a "trace without a crime"⁽⁵⁾:

An image of reality having all the characteristics of an imprint, without it really being an imprint of something, without, that is, such a reality existing or ever having existed: a pure product of synthesis. In its essence, the digital image can not only alter reality, like the analogue

(5) This issue is much more complex and nuanced than that. On the debate around the loss of indexicality in digital media, see Marra (2006).

image, but can generate it as Integral Reality or Hypereality (Gurisatti 2019, p. 330).⁽⁶⁾

According to Gurisatti, there are some turning points in the history of technology that are a real breakthrough. These ruptures made the transition to the synthetic image and the society of simulacra possible and inaugurated a new trend. A double rupture occurs first with the advent of photography and then with the advent of the digital. The first revolution occurs with the advent of photography and the technical reproducibility of the image. The object is detached from the here and now and becomes infinitely reproducible. The object is translated into a manipulable image that lends itself to new re-combinations (Gurisatti 2019, pp. 341–344),

The second turning point is the advent of the synthetic image, the detachment of the image from the object. The image becomes “de-realised” (Gurisatti 2012). The synthetic image is created *ex-nihilo* from existing image archives. Gurisatti (2019, pp. 334–338) discusses the ontological nature of synthetic images against the background of pictorial images and analog images. He sets out three main attributes of the synthetic image:

1. it is a “digitography” (from the English word *digit*, number) that is, a writing of numbers; it is synthetic in the sense that “the physical data do not leave a direct imprint, but are recorded, transformed, processed and represented in the form of electrical charges / signals” (Gurisatti 2019, p. 336);
2. it has a “virtual” origin in the sense of possibility and potentiality as it is discontinuous rather than discrete;
3. it has a simulation status both in the sense of imitation and representation as well as with the meaning of fiction, deceit and lying (Gurisatti 2019, p. 337).

(6) “Un’immagine della realtà avente tutte le caratteristiche dell’impronta, senza che essa sia davvero impronta di qualcosa, senza cioè che tale realtà esista o sia mai esistita: un puro prodotto di sintesi. Nella sua essenza l’immagine digitale non solo può *alterare* la realtà, come l’immagine analogica, ma può *generarla* come Realtà Integrale o Ipperrealtà” (Gurisatti 2019, p. 330).

4. Simulation and reality

These problems and in particular the creation of videos that blur the real/fake distinction calls into question the concepts of simulation, model, and representation. The simulation aspect of synthetic image is important. As Gianfranco Bettetini pointed out, “computer graphics seems to tell us that [...] each image can be reduced to algorithms to enter a computer program or, in any case, it can be transformed into a model that can be simulated mathematically” (Bettetini 1991, p. 10)⁽⁷⁾.

The notion of model as such is used in various disciplines, from mathematics to science and literature. It goes without saying that there is no agreement on what a model is or should be, nor is there a uniform use of the term in different fields of knowledge. Even considering the field of semiotics alone, we must come to the conclusion that this concept is used in quite different ways. For example, W. Nöth (2018) lists various definitions of this concept. Models are physical, conceptual or abstract representations of objects or situations. A map is a model; modeling simulation, mathematical and scientific models are other examples. A recurring aspect of the concept of model, however, is that the relationship between the model and the modeled object is such that between the two there is an asymmetrical relationship:

These images (models) tend to establish between themselves and the modeled objects a relationship such that all the elements and objects that (from the pragmatic angle of a user of a given model) appear in the modeled object, will also appear in the image (in the model), although the reverse case may not occur (Ivanov 1962, p. 183).

It is clear that, in the near future, the modeling capacity of virtual reality and artificial intelligence systems will be such as to make the model indistinguishable from the modeled object. Deepfake videos are a simulated reality and, as such, call into question the concept of simulation. As Renato Betti (1981, p. 947) writes, “to simulate means to

(7) All translations from Italian are my own: “la computer graphics sembra dirci che [...] ogni immagine può essere ridotta ad algoritmi per entrare nel programma di un calcolatore o, comunque, può essere trasformata in un modello simulabile matematicamente”.

give the appearance of something else. Simulating also means obtaining the effect of something else”⁽⁸⁾.

As Betti writes again:

Consider a model in its most evolved form of symbolic representation of a current situation, which can be used, in relation to some aspect, in place of the given situation. The replacement takes place for different reasons. For example, because the original system is inaccessible, like an astronomical system, or because its direct observation is extremely expensive. Or because it is too complex, like an economic or social system, which it is impossible to describe with a set of relations for which sufficient analytical calculation techniques have been developed. Or again because you want to exercise a sort of forecast, for example by altering the time scale and determining the behavior to be expected in the future. Through these three functions stated — reduction to increase material access, simplification, forecast — all the formal substitutes for a current situation pass (Betti 1981, p. 948).⁽⁹⁾

Some argue, however, that the question of the indistinguishability of the model from the modeled object is irrelevant from the point of view of the analysis of virtual reality, since there is no actual difference between virtual and real, both being related to reality, even if virtual reality does not physically exist. A similar thesis, for example, was recently put forward by the philosopher David Chalmers in his study *Reality +. Virtual Worlds and the Problems of Philosophy* (2022). Indeed, Chalmers’ basic thesis is that although the physical world and the virtual world are in some respects different, they are both real and what happens in simulated realities actually happens and is therefore real. Much,

(8) “simulare significa dare l’apparenza di qualcos’altro. Simulare significa anche ottenere l’effetto di qualcos’altro”.

(9) “Si consideri un modello nella sua forma più evoluta di rappresentazione simbolica di una situazione corrente, il quale può essere utilizzato, in relazione a qualche aspetto, al posto della data situazione. La sostituzione avviene per motivi diversi. Ad esempio perché il sistema originario è inaccessibile, come un sistema astronomico, o perché la sua osservazione diretta è estremamente costosa. Oppure perché è troppo complesso, come un sistema economico o sociale, che è impossibile descrivere con un insieme di relazioni per le quali siano sviluppate sufficienti tecniche analitiche di calcolo. Oppure ancora perché si vuole esercitare una sorta di previsione, alterando ad esempio la scala temporale e determinando il comportamento da aspettarsi nel futuro. Attraverso queste tre funzioni enunciate — riduzione per aumentare l’accesso materiale, semplificazione, previsione — passano tutti i sostituti formali di una situazione corrente”.

therefore, depends on the definition of the concept “real” that researchers assume in the study of these phenomena.

5. Grappling with deepfakes

This said, it is worth dwelling on a paradoxical situation going on in the field of semiotics. As it is well known to all the experts of this subject, there is a famous definition of semiotics that interlocks the study of signs with the possibility of lying by means of signs (Eco 1975). The story has so often been told that I am exonerated from repeating it once more (Danesi 2017; Gramigna 2020). One would expect that the natural development of semiotics would include the study of the semiosis of deception and that, broadly speaking, phenomena like fakes, illusions, simulations, deception and all devices of misrepresentation would appear to be the proper domain of semiotics. In fact, quite the contrary is true. Despite the fame of Eco’s definition, semiotic research on this subject has been scant and fragmentary, with a few rare exceptions. Save Eco’s formula, semiotic research has been addressing the subject of deception sparingly. As Danesi (2017, p. 20) pointed out, “it is somewhat surprising to find that virtually no one has approached sign analysis from Eco’s perspective, even though it goes way back to 1976”. This is quite paradoxical, to say the least. By this claim I do not mean to disregard the valuable, although non-systematic contributions to this field (Anderson 1986; Bonfantini et al. 1997; Caporale, Demaria, Donati, Lorusso, Mazzuchelli 2022; Danesi 2017; 2020; Eco 1987; 1997; Fiordo 1990; Fabbri 1991; 1998; 2010; Gramigna 2013; 2020; 2022; Jervis 1970; Nöth 1997; Nuessel 2013; Pelc 1992); it is rather an encouragement to catch up with such a pivotal issue. Recent studies dealt with the issue of disinformation and post-truth (Ferraris 2017; Lorusso 2018; Polidoro 2018) and the study of deepfakes from a semiotic perspective (Châtenet 2022; Dondero 2021; Leone 2022; Santangelo 2022)⁽¹⁰⁾. Also, within the domain of philosophy only a few papers devoted to deepfakes were published up to now (Fallis 2020; Floridi 2018; Kerner, Risse 2021; Öhman 2019; Rini 2020).

(10) On this issue, see also A. Baricco’s article “La verità sulla post-verità. Perché questa definizione è infondata” published in *Robinson di Repubblica* (April, 30, 2017).

Work in this field has been actively carried on by philosophers, psychologists, and sociologists. In the macro-field of deception studies covered by various disciplines and approaches, we can single out two distinct areas of research (Vincent-Marrelli 2004, p. 23). The first area, which can be called “methodological”, is typically found within philosophy and linguistic speculations — especially in pragmatics. The underlying theoretical question addressed in this area is this: ‘What are the underlying principles that permit understanding in interaction?’ In this sub-field of studies, Paul Grice’s conversational maxims and the principle of cooperation is key (Vincent-Marrelli 2004, p. 23). The second area of research focuses on truthfulness as truth-telling, where it is seen often in opposition to “falsehood”, “lying”, “deception”, and “deceit” (Vincent Marrelli 2004, p.23). In this second area, we find both ontological and deontological discussions. The focus is typically on the strategies and types of deception. As Ray Hyman (1989) suggests, given the heterogeneous nature of this phenomenon, a general theory of deception cannot but be interdisciplinary. However, psychology, philosophy, pragmatics, semiotics and other disciplines have given only fragmentary contributions, although very enlightening.

6. U. Eco’s semiotic theory of fakes

We owe to Umberto Eco the first formulation of a fully-fledged semiotic theory of falsification (Eco 1987; 1988; 1990)⁽¹¹⁾. Eco set up a very sophisticated typology of fakes, of which I will lay out the main categories that can be useful for the present study. The first type of Eco’s semiotics of falsification is doubles. Doubles concern the replicability of an object and they do not intent to deceive. A double is a physical occurrence that has got all the properties of another physical occurrence. Examples of doubles are two eggs, or two sheets of white paper which can be used for the same purpose. The second head in Eco’s typology is pseudo-doubles. There is a pseudo-double when one single occurrence takes on a particular value for one of the following reasons: because of its origin, as in this

(11) For a discussion on Eco’s take on falsehoods and forgeries see, Lorusso (2021) and Palazzolo (2018).

example of the first car Model T produced by Ford, or because the object was used in a particular context, as in the example of the Holy Grail used by Jesus Christ. Now we have reached the third category of falsification and the most important point for the purpose of this study: false identification. False identification may occur when there are two different objects — object A and Object B — produced by two different authors in two different historical contexts. These two objects can be identified as being identical or indiscernible either by an individual or by a group of individuals who decide that the two objects are identical. This leads to a host of different problems. First of all, the issue of intentionality. There are two different levels of intentionality at stake. Firstly, the intention of the people who produced object B. Secondly, the intention of the people who identified object B and object A as identical.

If we consider the people who produced object B, we can distinguish a *bona fide* and *mala fide* intention. The author B might have produced the object B maliciously to pass off something fake for real, but he might also have acted in good faith. More complex is the situation of the intention of those who decide that Object A and Object B are identical. A consideration of the subject's knowledge and beliefs adds complexity to the problem. According to the type of belief the individual or the group has, Eco (1988) singles out four different cases. The first case is a confusion between identity and interchangeability. Someone knows very well that Object A cannot be identified with Object B, which was produced afterwards and by imitation, but believes that the two objects are interchangeable in value and function and uses and presents one as identical to the other. The second case is the case of malicious false identification. Someone knows that Object B is just an imitation of Object A and cannot be identified with it, nor does he believe that the two objects are interchangeable. But, in bad faith, he pretends (and says) that O_b is identical to O_a . There can also be false identification because of inexperience of the individual. He ignores that the two objects are not identical. The fourth type is the presumption of interchangeability. Someone knows very well that O_a and O_b are physically different but decides that, under a certain description and for the purposes of a certain practice, one is as good as the other and is presented not as identical but as completely interchangeable.

7. Before deepfakes: digital manipulation and the falsification of reality in photography

The manipulation and falsification of images is not new. Image fakery has a long pedigree and there are plenty of examples in art history of fakes and forgeries (Brugioni 1999; Chéroux 2003; Farid 2003; Marra 2006; Nickell 1994; Smargiassi 2009;). Image fakery gained momentum with the rise of photography and today has reached its pinnacle through the use of artificial intelligence and deep learning techniques. The different methods of photo fakery have been brilliantly studied by Dino Brugioni (1999), who is an authority in this field. He pointed out four ways in which you can manipulate a photograph. First, information can be removed from an image. The second way of manipulating a photograph is by adding information to it. The third way is through photomontage. You take one or more photographs and you combine them. The fourth way is when you have a photograph which is true, but what it purports and depicts is false. This technique is called “false captioning” (Brugioni 1999, pp. 17–23). There is a key distinction which can be drawn from these examples. There are two broad categories that can be distinguished: “falsification” and “forgery”. Falsification occurs when an object has been altered in some ways or another to make it look different from what it is. A forgery, in the strict sense of the word, is an object made to be passed off as what is not — as for instance an object passed off as an antique.

8. Deepfakes culture

Within the larger area of the techniques of digital manipulation my interest lies in the technique called “deepfake”. Deepfakes videos are videos in which faces have been swapped using machine learning and deep neural networks, which is what the term “deep” in word “deepfake” stands for. The output is the creation of very realistic simulation of existing videos or audio recordings. To be more accurate, deepfakes videos emulate the style and the way in which authentic videos work. Not only deepfakes emulate and mimic authentic videos but they are also able to produce videos *ex nihilo*, by using data sets of images.

There are different varieties of deepfakes, and are produced for different purposes. The common denominator is that they mimic an existing content making the fake look like real. Deepfakes were introduced in 2017 and it was quite a jump in the world of digital manipulation. I believe that the blurring of the distinction between authentic and fake videos poses important epistemological questions that semiotics cannot ignore.

The term “deepfake” is fairly new and it lumps together the words *deep learning* (a set of techniques that enables artificial intelligence to learn forms) and *fake* (as for false or fake news). Deepfakes generally use face images or body images extrapolated online that are then adapted and used in a context different from the original one. Deepfakes provide the technical possibility to forge the image or video of a person purporting her saying or doing things as if she was the real person. The deepfake purporting former US President Barack Obama produced by BuzzFeed is already a milestone in this field. The Merriam Webster dictionary defines the term deepfake as “an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said”⁽¹²⁾ and the Cambridge Dictionary as “a video or sound recording that replaces someone’s face or voice with that of someone else, in a way that appears real”. The definitions, however, vary as there has been a growing amount of literature in the fields of law, computer science and media studies.

9. Concluding remarks

A preliminary question is whether deepfakes fall into the basket of forgery, falsification or a mix of both. This distinction ties very well with the main structure of deception as described by Bowyer J. Barton (1982). He singled out two main ways in which deception occurs: dissimulation and simulation. Whilst the purpose of dissimulation is to hide the real, the purpose of simulation is to display the false. Both dissimulation and simulation include three different techniques of

(12) <https://www.merriam-webster.com/dictionary/deepfake>.

deception which makes in total six basic ways of deceiving. For the time being, I'm interested in only two of those ways for they parallel to the two main concepts I laid out before, namely, falsification and forgery. In mimicking, a replica of reality is created by selecting one or more characteristics of the real in order to achieve an advantageous effect. A good example of mimicking is face swap. "Face Swap" is a technique through which the face of a famous politician has been changed with the face of another with a very realistic and uncanny effect. Different is the case of "inventing". In inventing the false is displayed through the fashioning of an alternative reality and not simply through the mimicking of the existing reality.

What are the implications that can be drawn from this discussion of deepfakes? Is the future lying ahead of us a dystopic scenario where fakery becomes harder and harder to identify? Seeing is not anymore believing. So, one question would be to lay out the critical skills people need to operate in this scenario. If fakes and authentic are indistinguishable to the average internet user there must be other criteria (such as content, context, historical records of the speaker, etc.) in order to establish whether something we see is authentic or not. Perhaps in the future there will be a system of authentication by the author himself, but this remains an open problem.

One of the most important implications from the semiotic point of view is that two different levels are often lumped together. The aspects of convincingness, truth and denotation of signs (may that be a speech, a visual sign or a video) are not the same although they are very often mixed. The result of mixing these levels together is adding confusion to an aspect which is already complex in and of itself. Being convincing does not equal to being truthful nor to being reliable. From the semiotic stance, deepfakes as well as any type of misinformation capitalizes on the mixing up of two aspects of the uses of signs.

The second point that can be drawn is the following. Whilst in the past fakes and forgeries were about falsifying art objects and artifacts, today's target of digital manipulation is the essence of the individual, the human face, the speech and his actions. In a nutshell, his or her total physiognomy, taken in its broad sense. What is considered to be the core of the identity of the individual — the face, the voice, and the mimic — is being faked. If the quantity of images circulating on the web is directly proportional to

the possibility of being the target of a deepfake, then almost everyone can be a target, not only celebrities and politicians. In the case of malicious false identification, this has considerable implications on privacy and ethics, especially where there is no consent from the people involved.

There is a third and the last point to consider. If we think of art history, we have a situation in which one single valuable item — conceived as a *unicum* — is the object of falsification by many skilled forgers. This means that one authentic object generates a host of fakes and copies that seek to mimic and pretend to be like the original. Today, with the emergence of deepfakes, it seems that the model of faking follows a reverse logic. There is not a one-to-one replica but a model of many-to-one. The production of deepfakes, in fact, entails the extraction of images of face through deep learning. This process requires a very large amount of data. In other words, in order to create one deepfake video you would need thousands of face pictures. This model of producing fakes is somewhat a reverse of the model of the past.

References

- ALBAHAR M., ALMALKI J. (2019), *Deepfakes: Threats and countermeasures systematic review*, “Journal of Theoretical and Applied Information Technology”, 97(22): 3242–3250.
- ANDERSON M. (1986), “Cultural concatenation of deceit and secrecy”, in R.W. MITCHELL, N.S. THOMPSON (a cura di), *Deception. Perspectives on Human and Nonhuman Deceit*, State University of New York Press, Albany, 323–348.
- BARICCO A. (2017), *La verità sulla post-verità. Perché questa definizione è infondata*, “Robinson. La Repubblica”, 30 Aprile: 45–47.
- BETTETINI G. (1997), *La simulazione visiva. Inganno, finzione, poesia e computer graphics*, Bompiani, Milano.
- BONFANTINI M.A., C. CASTELFRANCHI, A. MARTONE, I. POGGI, J. VINCENT (1997) (a cura di), *Menzogna e simulazione*, Napoli, Edizioni Scientifiche Italiane.
- BETTI R. (1981), “Simulazione”, in *Enciclopedia Einaudi. Vol. 12 Ricerca/Socializzazione*, Einaudi, Torino, 944–971.

- BAUDRILLARD J. (2004), *Il Patto di lucidità o l'intelligenza del Male*, tr. it. di A. SERRA, Raffaello Cortina, Milano.
- BOLINGER D. (1973), *Truth is a linguistic question*, "Language", 49(3): 539–550.
- BOWYER J. BARTON (1982), *Cheating. Deception in war & magic, games & sports, sex & religion, business & con games, politics & espionage, art & science*, St. Martin's Press, New York.
- BRUGIONI D. (1999), *Photo Fakery. The History and Techniques of Photographic Deception and Manipulation*, Brassey's, Dulles, Virginia.
- CAPORALE M., DEMARIA C., DONATI D., LORUSSO A.M., MAZZUCHELLI, F. (a cura di) (2022), *Le forme del falso*, Bologna University Press
- CASTELFRANCHI C., YAO–HUA T. (a cura di) (2011), *Trust and Deception in Virtual Societies*, Kluwer Academic Publishers, Dordrecht.
- CHALMERS D.J. (2022), *Reality +: Virtual Worlds and the Problems of Philosophy*, Allen Lane, UK.
- CHÂTENET L. (a cura di) (2022), *Images, mensonges et algorithmes: La sémiotique au défi du Deep Fake*, special issue of "Interfaces numériques", 11, 2 (26 July 2022). DOI: <https://doi.org/10.25965/interfaces-numeriques.4847>, Université de Limoges; Faculté des Lettres et Sciences Humaines.
- CHÉROUX C. (2003), *Fautographie: Petite histoire de l'erreur photographique*, Yellow Now, Crisnée, Belgium.
- DANESI M. (2017), "Eco's definition of semiotics as the discipline of lying", in T. THELLEFSEN E B. SØRENSEN (a cura di.), *Umberto Eco in His Own Words*. de Gruyter, Boston and Berlin, 19–25.
- DANESI M. (2020), *The Art of the Lie. How the manipulation of language affects our minds*, Prometheus Books, Guilford, Connecticut.
- DONDERO M. (2021), *Composition and Decomposition in Artistic Portraits, Scientific Photography, and Deep Fake Videos*, "Lexia", 37–38: 439–454.
- DESMET M. (2022), *The Psychology of Totalitarianism*, Chelsea Green, London
- ECO U. (1975) *Trattato di semiotica generale*, Bompiani, Milano.
- (1987), *Fake, identity and the real thing*, Special issue of "Versus", 46, Bompiani, Milano.
- (1988), "Tipologia della falsificazione", in *Monumenta Germaniae Historica Schriften. Teil I. Fäschungen Mittelalter*. Hansche Buchhandlung, Hannover.
- (1990), *I limiti dell'interpretazione*, Bompiani, Milano.

- (1997), “Dire il contrario”, in M.A. Bonfantini et al. (a cura di), *Menzogna e Simulazione*. Edizioni Scientifiche Italiane, Napoli, pp. 33–43.
- FABBRI P. (1991), *Appunti sul vero e sul falso*, “Sfera”, 20, Ed. Sigma-tau, Roma.
- . (1998), *La voce è matta*, “Versus. quaderni di studi semiotici”, 79: 9–26.
- . (2010), in C. Casarin e D. Fornari (a cura di). *Estetiche del camouflage*, Et. al. Edizioni, Milano, 8–20.
- FALLIS D. (2020), *The epistemic threat of deepfakes*, “Philosophy and Technology”, 34: 623–643.
- FARID H. (2003), *A picture tells a thousand lies*, “New Scientist”, 179 (2411): 38–41.
- FERRARIS M. (2017), *Postverità e altri enigmi*, Il Mulino, Bologna.
- FILIMOWICZ M. (2022), (ed.) *Deep Fakes. Algorithms and Society*, Routledge, London and New York.
- FIORDO R. (1990), *From sincerity to mendacity in personal rhetoric: A discrete look at continuous feelings*, “Semiotica”, 80 (1/2), 89–107.
- FLORIDI L. (2018), *Artificial Intelligence, Deepfakes and the Future of Ectypes*, “Philosophy and Technology”, 31: 317–321.
- GRAMIGNA R. (2013), *Augustine on lying: a theoretical framework for the study of types of falsehoods*, “Sign Systems Studies”, 41: 446–487.
- (2020), *Augustine’s Theory of Signs, Signification, and Lying*, De Gruyter, Berlin.
- (2022), *Faces in disguise. Masks, concealment, and deceit*, “Topoi”, <https://doi.org/10.1007/s11245-022-09806-0>.
- GRAMIGNA R., VOTO C. (2020), “Semiotica, prossemica e contagio. Il senso delle distanze ai tempi del Covid-19” in M. LEONE (a cura di), *Volti virali*, FACETS Digital press, Torino, 131–150.
- GURISATTI G. (2012), *Scacco alla realtà. Estetica e dialettica della derealizzazione mediatica*, Quodlibet Studio, Macerata.
- (2019), *Il delitto perfetto: Strategie dell’immagine tra analogico e digitale*, “Scenari: Quadrimestrale di approfondimento culturale”, 10 (1): 330–352.
- HYMAN R. (1989), *The psychology of deception*, “Annual Review of Psychology”, 40: 133–154.
- IVANOV V. (1962), “Premessa alle tesi del Simposio sullo studio strutturale dei sistemi di segni”, in C. Prevignano (a cura di), *La semiotica nei paesi slavi. Programmi, Problemi, Analisi*, Bompiani, Milano, 181–185.

- JERVIS R. (1970), *The logic of images in international relations*, Princeton University Press, Princeton.
- KARNOUSKOS S. (2020), *Artificial Intelligence in digital media: The era of Deepfakes*, "IEEE. Transactions on Technology and Society", 1(3): 138–147.
- KERNER, C. & RISSE, M. (2021), *Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds*, "Moral Philosophy and Politics", 8 (10) 81–108.
- LEONE M. (2022), *L'idéologie sémiotique des deepfakes*, "Interfaces numériques. Images, mensonges et algorithms: La sémiotique au défi du Deep Fake", special issue, 11 (2): 1–16.. DOI: <https://doi.org/10.25965/interfaces-numeriques.4847>.
- LORUSSO A. (2018), *Post-verità. Fra Reality tv, Social Media e Storytelling*, Laterza, Bari.
- (2021), *Power, dangers and resources of forgery, from theory to novels*, "Rivista di Estetica", 76: 95–112 <https://doi.org/10.4000/estetica.7690>.
- MATULEWSKA A., WAGNER A., MARUSEK S. (2022), *Masked covid life: A socio-semiotic investigation*, "Semiotica", 247: 55–85. <https://doi.org/10.1515/sem-2021-0035>.
- MESKYS E., LIAUDANSKAS A., KALPOKIENE J., JURCYS P. (2020), *Regulating deep fakes: Legal and ethical considerations*, "Journal of Intellectual Property Law & Practice", 15(1): 24–31.
- NICKELL J. (1994), *Camera Clues: A Handbook for Photographic Investigation*, The University Press of Kentucky, Lexington.
- ÖHMAN, C. (2019), *Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography*, "Ethics and Information Technology", 22: 133–140.
- FONTCUBERTA J. (2018), *La furia delle immagini. Note sulla postfotografia*, Einaudi, Torino.
- GREIMAS A. (1989), *The veridiction contract*, "New Literary History", 20 (3): 651–660.
- MARRA C. (2006), *L'immagine infedele. La falsa rivoluzione della fotografia digitale*, Mondadori, Milano.
- NUESSEL F. (2013), "Lying — A semiotic perspective", in J. PELKEY E L. SBROCCHI (a cura di), *Yearbook of the Semiotic Society of America*. Ottawa: Legas Press, pp. 151–164.

- NÖTH, W. (1997), "Can pictures lie?", in W. NÖTH (a cura di.), *Semiotics of the Media. State of the Art, Projects, and Perspectives*, de Gruyter, Berlin and New York, 133–146.
- (2018), *The semiotics of models*, "Sign Systems Studies", 46(1): 7–43
- PALAZZOLO G. (2018) *Eco e i dispositivi di falsificazione*, "Siculorum Gymnasium", LXXI (IV): 271–290.
- PELC, J. (1992), "Lie versus truth, falsehood and fiction from the semiotic viewpoint", in G. MARRONE (a cura di), *La Menzogna*, Quaderni del Circolo Semiotico Siciliano 34–35, Palermo, 243–249.
- POLIDORO P. (ed.) (2018), *Fake News, Misinformation / Disinformation. Special Issue*, "Versus. Quaderni di studi semiotici".
- PARISER E. (2011), *The Filter Bubble. What the Internet is Hiding from you*, Penguin, New York.
- POULSEN S.V. (2021), *Face-off — a semiotic technology study of software for making deepfakes*, "Sign Systems Studies", 49 (3–4), 489–508.
- RINI, R. (2020), *Deepfakes and the Epistemic Backstop*, "Philosophers' Imprint", 20 (24): 1–16.
- SANTANGELO A. (2022), "Il volto del futuro nell'era dei deep fake", in M. LEONE (a cura di) *Il metavolto*, FACETS digital press, Torino, 19–41.
- RUESCH J. (1972), *Semiotic Approaches to Human Relations*, Mouton, The Hague, Paris.
- SMARGIASSI M. (2009), *Un'autentica bugia. La fotografia, il vero, il falso*, Contrasto, Roma.
- SUNSTEIN C.R. (2007), *Republic.com 2.0*, Princeton University Press, Princeton and Oxford.
- VACCARI C., A. CHADWICK (2020), *Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news*, "Social Media + Society", 6: 1–13.
- VINCENT-MARRELLI J. (2004), *Words in the Way of Truth. Truthfulness. Deception and Lying Across Cultures and Disciplines*, Edizioni Scientifiche Italiane, Naples.

MITO E REALTÀ DELL'IA: UNO SGUARDO SEMIOTICO

GUIDO FERRARO*

ENGLISH TITLE: *Myth and reality of AI: a semiotic perspective*

ABSTRACT: Looking from a specifically semiotic point of view, several issues about AI have for us a special interest. We should consider both the aspects on which semiotics could offer its own specific contribution, and those for which, conversely, it can receive, from AI studies, important stimuli and theoretical inspirations. Mainly, this is perhaps an opportunity to focus on some decisive aspects of semiotic theory, so as to updating them, taking also into account those new dimensions of cultural reality.

Among the main points: the relationship between the myth-making about AI and its actual capacities of practical contribution, the relationship between machine intelligence and human intelligence, the questions about different forms of creativity, the methods for classifying and labelling cultural products. But mainly the key questions concern the possibilities of a software to grasp that fundamental semiotic entity that is the meaning, and therefore the issue of intersubjectivity, and ultimately the inescapable specificity that characterizes us, as human beings.

KEYWORDS: Semiotics, Artificial Intelligence, Creativity, Tagging, Human nature.

I. L'intelligenza artificiale in prospettiva semiotica

Credo che sia opportuno definire innanzi tutto la prospettiva in cui questo articolo è scritto: per sede di pubblicazione e area disciplinare di pertinenza, nonché per le mie competenze personali, si tratta essenzialmente della prospettiva degli studi semiotici. Certo non vi

* Università degli Studi di Torino.

è bisogno di sottolineare quante dimensioni d'interazione, significative e affascinanti, vi siano tra i nostri studi e gli sviluppi dell'IA, e quanti siano i temi sui quali la semiotica potrebbe essere in grado di offrire un contributo alle riflessioni degli specialisti di IA — cosa che però, constatiamo con rammarico, accade troppo di rado, e su questo dobbiamo riflettere. Dall'altro lato, vi sono anche molti aspetti per i quali sviluppi e prospettive relativi all'IA possono a loro volta offrire alla semiotica spunti di riflessione anche molto importanti, e questo lato è per me forse il più rilevante. Del resto, molto di quello che ho cercato di fare, proponendo per la nostra disciplina linee di sviluppo più attuali e innovative — quella che un po' ironicamente ho chiamato *Semiotica 3.0* (v. Ferraro 2019) — è stato di trovare una sintesi più ricca e avanzata tra gli insegnamenti dei nostri maestri, introducendo in tale quadro anche i contributi che possono appunto venire da altre discipline. È proprio in questa linea, dunque, che penso sia utile far tesoro delle indicazioni e degli stimoli che possiamo trarre dagli studi sull'Intelligenza Artificiale.

Va però anche ricordato che uno dei compiti propri alla semiotica consiste nell'analisi dei sistemi di pensiero e di rappresentazione, mettendone in luce grammatiche interne e modalità di costruzione di senso. Al di là dei suoi aspetti tecnici e delle funzionalità delle sue applicazioni, il discorso sull'IA è senza dubbio anche ricco di dimensioni simboliche e di valenze ideologiche, tanto da occupare un posto di rilievo nell'immaginario contemporaneo, e dunque anche di questo vorrei tener conto. Tanto più che, a rifletterci, in prospettiva semiotica uno degli aspetti più significativi potrebbe consistere proprio nel fatto che la linea d'evoluzione dell'IA presenta un'inconsueta ambivalenza per la quale, a ben vedere, tanto l'immaginario fantascientifico si ispira agli avanzamenti tecnologici quanto questi ultimi prendono spunto, sempre più decisamente, dalle proposte dell'immaginario. Ariel Kyrrou (2020), che sottolinea questa speciale peculiarità, la collega ad un'altra, non meno singolare, ambivalenza: perché l'IA ci si presenta al tempo stesso nei termini di una delle più grandi meraviglie della storia umana e come una comoda aiutante nelle piccole faccende quotidiane. Sembrerebbe trattarsi di due dimensioni completamente diverse e irrelate, ma è invece forse corretto vederle come intimamente intrecciate.

La novità è proprio questa: è la vostra segretaria così terribilmente precisa, la vostra donna delle pulizie così servizievole, il vostro autista così esperto e puntuale... sono loro a condurvi inaspettatamente in un mondo completamente nuovo, che vale come una fantastica, radicale reinvenzione dei vostri modelli di vita. Vi fiderete di loro, come non vi sareste fidati di altri.

In ogni caso, una prima prospettiva d'analisi dovrebbe essere senz'altro quella che definisce semioticamente i modi in cui molti scritti sull'IA, con artifici retorici talvolta anche banali ma comunque evidentemente efficaci, ne propongono una visione *mitizzante*, giocata sull'immagine di un futuro dal sapore fantascientifico, ma alla vecchia maniera: un futuro brillante quanto affascinante, senz'altro disomogeneo rispetto al mondo attuale... Ma, bisognerebbe forse dire, disomogeneo rispetto al *mondo reale*, dal momento che, come tutti sappiamo, dispositivi basati su qualche forma di IA sono presenti da anni, parte di una nostra — non così brillante ed elettrizzante — realtà quotidiana (anche per questo, ho scelto di preferire poi in questa sede una riflessione su applicazioni, appunto, più attuali e quotidiane).

Questo atteggiamento mitizzante vorrebbe tra l'altro far dimenticare che l'IA ha una sua storia definita, che ne colloca l'origine in un preciso, e non così recente, momento storico, e ne inquadra i tratti di base in uno specifico clima culturale. Molti dei suoi riferimenti di partenza, legati alla cibernetica e alla teoria dell'informazione degli anni quaranta e cinquanta del novecento, pur potendo apparire oggi per certi versi gravemente sorpassati, continuano ad avere un peso innegabile anche negli studi più attuali — e su questo la semiotica avrebbe certo delle cose da dire. Lungi dal costituire una novità incontaminata e quasi incorporea, quello dell'IA è un progetto concepito da persone con una definita formazione e un definito orientamento. L'IA di cui oggi parliamo è parte di un apparato industriale, si fonda su un insieme di definite tecnologie materiali, fa parte tra l'altro anche dei piani di rinnovamento delle basi dell'economia mondiale promossi da un determinato insieme di grandi gruppi finanziari: si tratta insomma di un progetto d'innovazione tecnologica che — pur se questo può essere in certa misura marginale in uno studio semiotico — ha molto a che vedere con questioni relative a nuovi assetti socioeconomici, rapporti di potere, meccanismi di controllo del comportamento di miliardi di persone.

È proprio in questo quadro che è importante sottolineare che, al di fuori del mito, applicazioni più banali di IA sono in uso da tempo, e che si tratta di una direzione di sviluppo di quegli stessi strumenti informatici che ormai ci sono per molti versi familiari, tanto in termini positivi, di utilità e miglioramento nei nostri modi di vivere e lavorare, quanto in termini negativi, vale a dire nei termini che hanno fatto delle applicazioni informatiche un proverbiale sinonimo di instabilità, incertezza di funzionamento e incostanza di risultati. A fronte della rappresentazione mitizzata di cui s'è detto, nell'esperienza reale la percezione diffusa resta quella di un'informatizzazione che esibisce anti-usabilità, che si avvale di *chatbot* ridicoli e irritanti, che viene a complicare ottusamente i rapporti dei cittadini con le istituzioni, che genera troppo spesso disorientamento e senso di insicurezza. E questa appare come una linea di tendenza in peggioramento, tanto che vale per molti quanto suggerito dalla constatazione esemplare di Samantha Cristoforetti, astronauta italiana per eccellenza: con la tecnologia di cinquant'anni fa velivoli con equipaggio umano riuscivano ripetutamente a fare il tragitto terra-luna e ritorno, mentre con la tecnologia attuale di fatto non siamo in grado di portare equipaggi umani oltre la cosiddetta "orbita bassa" intorno alla Terra. Così, si è fatta netta nell'immaginario diffuso — complice anche la cinematografia di fantascienza — un'opposizione simbolica tra la scintillante e affidabile tecnologia dell'epoca della *modernità* e gli impacci sopravvenuti in epoca di *postmodernità*, con una tecnologia che appare più debole perché più complicata e pretenziosa, volendo affidare alle macchine anche ciò che, parrebbe, sarebbe più semplice e sicuro lasciar fare agli umani. Di qui i dubbi sull'IA: come convincere le persone comuni ad affidare a nuovi sviluppi degli strumenti informatici la cura della loro salute o quella dei loro interessi economici, per non dire la gestione di uno scenario di guerra, o l'indirizzo dell'economia e della politica mondiale?

Ed ecco che, di fronte a queste difficoltà d'immagine, reali e in qualche misura giustificate, si cerca di rispondere, sul piano dei sistemi di rappresentazione, con la costruzione appunto di un mito che, con l'etichetta ammaliante di "Intelligenza Artificiale", intenderebbe suggerire la nascita di un'entità sostanzialmente nuova, capace di ogni cosa e in prospettiva quasi infallibile. Si delineano qui vari problemi, e insieme

varie angolature secondo le quali affrontarli, che corrispondono a temi sui quali la semiotica dovrebbe essere chiamata a intervenire, facendo chiarezza con l'impiego dei suoi specifici strumenti, e anche grazie all'esperienza acquisita nello studio dei modi in cui si organizza l'allestimento sistematico delle rappresentazioni ideologiche. Rappresentazioni ideologiche che spesso, sappiamo, giocano su forme di ambiguità e duplicità. Perché sappiamo anche che oggi l'immagine dell'IA è in certa misura offuscata dai vari modi in cui questa può essere implicata in forme di censura (particolarmente sentite quelle che concernono la censura in rete) e modalità di indesiderato controllo dell'azione e del pensiero individuale. Di qui il possibile senso dell'ambiguità rilevata da Ariel Kyrou, grazie alla quale si gioca sull'immagine di un'IA che indebitamente mischia tratti clamorosamente rivoluzionari con toni servizievoli e rassicuranti.

Ma, a proposito delle dimensioni d'interazione tra semiotica e IA, ne va aggiunta anche una più decisamente teorica, e assolutamente seria: poiché la speculazione sui concetti chiave dell'IA genera interrogazioni sulla natura stessa dell'intelligenza in quanto tale, su cosa davvero voglia dire "ragionare" e "inventare", "apprendere" e "essere creativi", "produrre senso", esprimere "consapevolezza", attivare "intersoggettività", "provare emozioni", e financo, alla fine, cosa voglia dire davvero "essere umani". E qui la semiotica non vede soltanto la possibilità di assumere un qualche suo *meta-sguardo*, analizzando dei testi rilevanti per definirne le strutture di senso: giacché quelli che vengono posti sono problemi, in molti casi, anche propriamente e immediatamente semiotici: la semiotica, io credo, offre una delle prospettive teoriche più decisive per affrontare la questione chiave del "cosa definisce la nostra qualità di esseri umani".

2. Oggetti "intelligenti"

Le applicazioni di IA sembrano presentarsi prevalentemente in due forme: come macchine dotate di capacità operative sofisticate, oppure come software dotati di speciali abilità. Siccome ovviamente anche le macchine sono fondate su software e processori di calcolo, la differenza tra le due

forme resta da precisare in altro modo. Volendo muovere da un oggetto specifico del tutto usuale, e direi tranquillamente, se non *banalmente* “intelligente”, assumo come primo esempio quello che personalmente mi ha dato in effetti l’occasione del primo contatto significativo con un oggetto intelligente: parlo della mia prima macchina fotografica, acquistata alla fine degli anni settanta (oltre cinquant’anni fa, dunque), dotata di un computer interno che le attribuiva la capacità di scegliere autonomamente le impostazioni d’esposizione (si trattava della allora avanzatissima Canon A1). Un libro, acquistabile a parte, ne illustrava funzioni e filosofia di fondo. Vi era spiegato che l’automatismo d’esposizione agiva grosso modo così: tramite un apposito sensore, la macchina leggeva i valori di luminosità di quanto veniva inquadrato, riferendolo ai vari settori componenti una griglia rettangolare predefinita. Otteneva così una mappa a zone delle luminosità, che passava poi a confrontare con un certo numero di analoghe mappe-modello (*types*) inserite all’origine nella sua memoria, e su questa base arrivava a decidere quale di tali modelli fosse più pertinente nel caso attuale (*token*). La macchina otteneva così l’indicazione di una coppia diaframma/tempo di scatto che, nel caso l’utente avesse optato per l’automatismo completo, veniva direttamente, dunque *autonomamente*, impostata sui comandi della macchina.

Nella sua semplicità, questo modo di procedere illustra un caso tipico, importante e diffuso, nel modo di funzionare di apparecchi “intelligenti”. La macchina, priva di capacità di apprendimento proprie, è dotata di *sensori* che le consentono la lettura di alcune proprietà del *mondo esterno*. È dotata inoltre di un *database* di conoscenze interne e di un *calcolatore* che le permette di confrontare i dati provenienti dall’esterno con i *types* presenti nella memoria interna; infine possiede degli *organi effettori* grazie ai quali può attivare determinati dispositivi o regolazioni. Interessante, ma tutto questo non la rende affatto simile a un essere umano, come qualcuno vorrebbe: non si tratta in effetti di un modo di procedere specificamente umano, ma del generico modello di un’azione volta a raggiungere uno scopo. Questo modello vale per gli esseri umani come per i pipistrelli, per i robot come per i marziani. C’è chi si esalta a dire che operazioni del genere coinvolgono dimensioni “psicologiche” (percezione, ragionamento, e così via), ma uno sguardo disincantato non vi coglie in effetti parvenze d’umanità.

In termini di modellazione teorica — e qui la pertinenza semiotica comincia a diventare evidente —, non è da trascurare il fatto che la procedura è volta a riportare il *caso osservato* a forme *precedentemente costituite*: si tratta dunque, in altre parole, di *riportare il nuovo al già noto*, e insieme i *dati osservativi* a *classi astratte definite all'origine*, o ancora il *contatto con il mondo reale* a una *gamma chiusa* di possibili combinazioni più astratte. Ci si chiede ovviamente da dove vengano i modelli dati alla macchina come predefiniti. Nel libro di cui dicevo, destinato a spiegare basi di funzionamento e filosofia operativa di quell'apparecchio fotografico, si spiegava che i modelli di cui era dotata la ROM della macchina erano stati definiti all'origine sulla base di un lavoro di raccolta preparatorio — procedimento in effetti del tutto comune nell'allestimento di molti tipi di oggetti “intelligenti”. Nel caso, in questo insieme di modelli si era formalizzato il comportamento effettivo di un certo numero di esperti fotografi, osservato e registrato in differenti situazioni. I *types* cui la macchina fa riferimento sintetizzano dunque uno specifico *saper fare* umano, opportunamente modellizzato e digitalizzato. L'intelligenza, nel caso, ha poco di “artificiale”: in pratica, è un po' come se la macchina avesse appreso un manuale ove fossero spiegate le scelte che un bravo fotografo opera nell'una o nell'altra situazione tipo.

Faccio il paragone con un manuale perché voglio sottolineare l'analogia — e poi le differenze — con l'insegnamento che può essere tratto da un *libro*, in quanto eminente contenitore di sapere. E si noti che “sapere” non è sempre soltanto una collezione di informazioni, ma anche un insieme di conoscenze concernenti tecniche operative, quel tipo di conoscenza che in semiotica chiamiamo “saper fare”. Ovviamente non sono solo i libri a valere come contenitori di intelligenza umana; si è detto anzi che questo vale parimenti per qualsiasi strumento. Entità materiali artificiali concepite per contenere intelligenza umana esistono dunque da sempre; sembra di poter dire, anzi, che esistano da prima della stessa origine del linguaggio, di cui potrebbero esser state a fondamento (cfr. Tomlinson 2015). Noi consideriamo giustamente un libro come uno strumento che sposta la nostra attenzione su dimensioni diverse dal qui-ora, uno strumento che descrive procedimenti possibili, che ci guida alla formulazione di ipotesi, che propone scenari immaginari... Ma qualcosa del genere vale per strumenti anche infinitamente

più semplici; nel suo piccolo, anche una pietra opportunamente lavorata suggerisce modalità di prensione e operazioni possibili, presenti anche solo a un livello virtuale, “cose che con questo strumento si potrebbero fare”, tramandando così tecniche e abitudini culturali. Noi siamo letteralmente circondati da oggetti che contengono intelligenza umana, forme di conoscenze e di saper-fare — il che pone certi aspetti dell’IA in una significativa linea di continuità che attraversa l’intera storia umana.

Ma cosa dunque rende diverso da questi l’odierno oggetto “intelligente”? A conti fatti, la sua capacità di passare dal *saper fare* al *fare*. Come nel caso della macchina fotografica: quando l’utente l’avesse voluto, non aveva che da inquadrare la scena, mettere a fuoco (oggi anche questo potrebbe essere automatizzato) e premere il pulsante di scatto. Nel caso della macchina presa ad esempio, questo vuol dire che, in concreto, il computer interno faceva in modo che le due tendine fisiche dell’otturatore scattassero con il tempo prescelto, che un apposito perno si spingesse nell’obiettivo a chiudere le lamelle del diaframma al valore precalcolato, eccetera: operazioni meccaniche, di cui è mirabile la rapidità e la precisione, ma che sono in sé prive d’intelligenza. Se dunque in tutto questo procedimento l’*intelligenza* è puramente umana, la macchina di suo ci mette questa sorta di istantanea e accuratissima capacità d’azione, e questa sì ci appare meno umana, e specifica anzi di una macchina — ma a pensarci, non è che questa capacità è in definitiva resa possibile proprio dal suo essere legata all’assenza di un’attività di riflessione e di scelta intelligente?

La prova di quanto ho detto possiamo vederla nel fatto che — qui il gioco si complica — gli esseri umani, da questi meccanismi d’automazione, tendono spesso a mantenersi a distanza di sicurezza. Ragionevolmente, cioè appunto: perché *ragionano*. E non è per un atteggiamento antitecnologico, per certi versi può essere anzi il contrario. È un fatto che la maggior parte degli utenti di apparecchi fotografici evoluti non si affida a un completo automatismo d’esposizione se non magari nel caso, appunto, in cui gli manchi *il tempo minimo per ragionare*, e così formulare una propria decisione. Ma perché non affidarsi alla macchina, se è vero che l’intelligenza della macchina in pratica sbaglia piuttosto di rado? Perché la macchina, siamo al punto decisivo,

non coglie la differenza tra il “non sbagliare” e il “fare bene”. Il carattere meccanico e istantaneo dell’intelligenza di una macchina raggiunge innegabili successi laddove vi è da calcolare un risultato “esatto”, mentre tutto diventa più discutibile quando ci si trovi di fronte alla *scelta* tra possibilità diverse tutte valide, che vengono a configurarsi quindi come scelte di carattere “espressivo”, come vale appunto per il nocciolo della competenza che un essere umano pone in atto nel realizzare una fotografia. Che valore possiedono procedure e modelli predefiniti in situazioni in cui, com’è facile osservare, ogni soggetto umano può fare scelte diverse, e addirittura la stessa persona in situazioni analoghe, in due momenti diversi ancora fa scelte diverse? Quello che sostanzialmente manca alla macchina non è un generico saper fare, ma un sapere più raffinato, relativo alla *intenzionalità di senso*.

In effetti, se immaginiamo di esplorare il “pensiero” di una macchina fotografica “intelligente” ci rendiamo conto del motivo dei suoi limiti, esprimibile di nuovo in termini prettamente semiotici: la macchina, in pratica, concepisce la fotografia come una mera *impronta del reale*. Potremmo dire, quasi, che le macchine dotate di questo tipo di “intelligenza” siano dotate anche di una specifica visione filosofica, illustrata come tale in tanti libri ispirati a una più o meno polverosa filosofia analitica — un tipo di pensiero che ancora ci riporta, non a caso, agli anni intorno alla seconda guerra mondiale. Che non si tratti di una mera questione tecnica bensì di un problema teorico di fondo, lo dimostra il fatto che la stessa difficoltà è rilevabile non solo in questo tipo di “oggetti intelligenti” ma anche negli atteggiamenti degli studiosi che, all’interno di diverse scienze umane, si sono maggiormente accostati a questo genere di prospettive. Si pensi ad esempio a quanti si richiamano a visioni cognitive in ambito di studi narratologici: li vediamo impiegare modelli teorici, ai nostri occhi eccessivamente semplificati, che sono fondati in primo luogo sul principio per cui, di nuovo, un racconto è una traccia, o un modello, di eventi che si svolgono nel reale, piuttosto che un modo per elaborare valori e costruire senso.

Se vogliamo guardare avanti e considerare quale contributo effettivamente la semiotica potrebbe fornire, dobbiamo ammettere però che la stessa semiotica (insieme tra l’altro alla vicina linguistica) è stata anch’essa fortemente toccata da questo tipo di inclinazioni. Vista con gli

occhi di adesso, la semiotica cresciuta tra gli anni quaranta e gli anni settanta del secolo scorso, per quanto fertile e innovativa, ci appare viziata da uno schematismo e da un oggettivismo semplificanti, accompagnati da un formalismo la cui superficie matematizzante valeva soprattutto a nascondere le effettive semplificazioni di fondo. Questi modelli possono mantenere oggi la loro utilità solo a patto di integrarvi una visione della *complessità* dei fatti semiotici (v. Ferraro 2021b).

3. Regole e creatività

In tema di complessità dei fatti semiotici, va affrontato un aspetto importante, che ci consente a questo punto di approfondire meglio quanto dicevo: le macchine possono molto bene applicare regole e modelli precostituiti, ma ben più difficilmente sanno elaborarne sviluppi creativi o innovazioni significative. È una questione molto discussa, e senz'altro interessante. Alcuni la vedono in modo banale: l'IA può solo applicare delle regole, ma non può seguire gli esseri umani nei modi in cui questi disattendono le regole, le cambiano, creano interi sistemi di regole nuovi, in altri termini fanno cose che sembrano stare a cavallo tra la devianza e la creatività, ma che in effetti per gli esseri umani sono comportamenti diffusi e del tutto normali. E la semiotica? La nostra scienza possiede in proposito indicazioni interessanti, ma non una teoria sistematica: tant'è vero che la maggior parte dei semiotici non sembrano in grado di definire una grammatica altro che come “un insieme di regole”, e si sentono in difficoltà quando si cita una nozione come quella di “codice” — nozione che in effetti, non aggiornata per decenni, esibisce un'aria piuttosto *vintage* (meglio prenderla con ironia, perché se pensiamo che si tratta di concetti chiave della semiotica, c'è da mettersi le mani nei capelli).

Intanto, potrebbe essere una buona cosa ispirarsi anche a certe indicazioni interessanti provenienti dal campo informatico, dove si è da tempo superata una visione normativa delle regole, a favore di concezioni fondate sulla contestualità e sulle *affordances*, sui rapporti eventi/processi concepiti in chiave di connessioni virtuali e di linee temporalmente parallele. Accettiamo senza problemi che ci sia qualcosa da

imparare, a livello teorico, dal modo in cui funzionano le macchine. Ma poiché questo non sembra poter condurre a forme “artificiali” di creatività, può essere forse più rilevante ciò che possiamo apprendere studiando il modo in cui soggetti umani particolarmente creativi riescono a produrre innovazioni, pur senza dar luogo ad alcun effetto di *sgrammaticatura*⁽¹⁾.

Innanzitutto, evitiamo di pensare alle grammatiche come *liste di regole*, giacché è molto meglio intenderle come insiemi di *risorse*. Per un semiotico, a pensarci, dovrebbe essere ovvio: una grammatica, più che dirti cosa puoi fare e cosa no, è lì per dirti che c'è un modo definito per esprimere X e un altro modo definito per esprimere Y. Ma la mia conclusione, studiando operazioni creative colte nel loro concreto modo di funzionare, è che per spiegarne il funzionamento dobbiamo immaginare le grammatiche secondo un modello di carattere *generativo*, dunque organizzato per strati gerarchici. Le formule, o se vogliamo “regole” più superficiali, possono essere comprese solo come espansioni, o specificazioni locali, di principi più generali e fondamentali, collocati su uno strato più profondo, e così per più livelli, fino ad avvicinarsi a un nucleo di principi di base di carattere primario e di portata molto generale. Per darne un'idea facendo un esempio del tutto elementare, un pittore creativo, poniamo ad esempio Matisse, ai livelli più fondamentali non viola alcuna regola di base sul funzionamento della pittura; usa invece risorse offerte dal sistema (impiego dei colori e dei modi di strutturazione topologica) per combinarle in forme che si discostano palesemente da altre opzioni possibili, quelle fondate per esempio sull'impiego della prospettiva. Ma il sistema della pittura mette a disposizione *più possibilità*, prevedendo valenze espressive diverse legate all'impiego dell'una o dell'altra di queste. Ci sono sempre più risorse di quante ciascun autore ne impieghi, ci sono più combinazioni possibili, di modo che possono aprirsi sempre strade differenti da quelle che altri hanno già percorso. L'artista che ci appare deviare rispetto a certe convenzioni di livello superficiale, sta aprendo una strada certo originale, ma lo fa attivando risorse che, a un livello gerarchicamente superiore, consentono appunto di congegnare alternative diverse. Se pensiamo un sistema espressivo

(1) Cfr. le mie considerazioni sull'innovazione, soprattutto in ambito musicale, in Ferraro (2022).

come una struttura gerarchica, possiamo rappresentarci le diverse specifiche varianti di linguaggio come lo sviluppo di opzioni disponibili a un livello superiore.

Mi fermo su questo modo di concepire la creatività perché un'idea in qualche modo analoga, l'idea di un sistema di principi operativi organizzato per livelli gerarchizzati, è stato significativamente introdotto nel caso delle macchine capaci di *deep learning*, dunque si tratta di una strada che sarebbe interessante percorrere in parallelo. Seguendo principi di questo tipo, una macchina "più intelligente" è capace di apprendere, modificando il proprio modo di agire e tenendo conto del contesto e delle interazioni in cui incorre, così da poter ridefinire le procedure prestabilite, pur se non gli obiettivi da raggiungere. Se questa sia o meno "creatività", dipende soltanto dal modo in cui decidiamo di definire tale concetto. Ciò che in ogni caso rende interessanti le macchine "che imparano" è che non si tratta di dire loro cosa debbono fare, ma di fornire un *obiettivo* rispetto al quale la macchina può formulare *proposte*, considerando linee operative tra loro alternative, e qui è ben possibile che si aprano strade che un essere umano non avrebbe considerato. Tuttavia, constatiamo che al centro di tutto resta non l'operatività della macchina ma l'intenzionalità, che è necessariamente umana, dato che la prima deve comunque discendere dalla seconda.

Perché questo è il punto decisivo. Come altri hanno osservato, in molti casi il supporto che le macchine possono fornire alla creatività è legato alla loro forza bruta di calcolo, che ad esempio consente loro di produrre un gran numero di combinazioni possibili, per poi tipicamente attenderne la valutazione da parte di un soggetto umano, o per fare confronti con valutazioni umane precedenti. La differenza fondamentale che rende diversa la creatività umana non sta in una qualche capacità in assoluto superiore ma in una condizione di fondo: la creatività umana è *orientata*. Tendiamo a non essere consapevoli della presenza e dell'importanza di questo carattere, pensando troppo spesso nei termini di una mera *innovazione*. Ma non dovrebbe essere questo il modo di vedere della semiotica. Quando degli artisti (scrittori, pittori, musicisti...) ci appaiono impegnati nel "cambiare le regole della grammatica", lo fanno non meramente per creare innovazione bensì in riferimento a un *mutamento definito* nei modi d'interpretare e rappresentare il reale,

dunque collocandosi all'interno di un definito *processo di trasformazione culturale*. Per richiamare un esempio, la mia analisi⁽²⁾ del celebre racconto *Deux amis* di Maupassant, oggetto di un ampio studio da parte di Greimas (1976), spiega come quel testo sia comprensibile solo come definita presa di posizione di un autore che nega un sistema culturale passato (grazie all'immagine paradossale dei due folli pescatori), per sostenere invece una nuova forma di cultura, che diremmo *moderna*. Maupassant non sta dunque meramente “innovando”, ma sta lavorando a una *trasformazione orientata* delle strutture narrative, allo scopo di adeguarle a un diverso sistema culturale. Se la creazione umana è dunque fundamentalmente orientata, non ha nulla di *neutrale e oggettivo*. E questo è un aspetto cui si dovrebbe dare grande attenzione, anche quando si pensa a processi creativi gestibili da un software d'IA.

4. La questione del senso

Ancora una volta, il problema è quello dell'accesso al senso. Alla fine, è chiaro che, per dar vita a forme più avanzate di IA, bisognerà che in certo modo, pur limitato e parziale, le macchine siano dotate di una qualche capacità di avvicinarsi a questa dimensione, come sappiamo del tutto fondamentale per gli esseri umani. Se ne sono ben resi conto molti di coloro che lavorano sull'IA, convinti che, se una macchina potesse passare dal livello del compiere operazioni a quello del cogliere significati, il salto sarebbe assolutamente decisivo. Il problema è che, ancora una volta, molti teorici dell'IA ci appaiono fermi a concezioni dell'aver senso piuttosto limitate.

Qui, diremmo più che in ogni altro ambito, la semiotica dovrebbe mostrarsi pronta a venire in aiuto. Ma è davvero così? In verità, disponiamo in proposito di un buon numero di testi interessanti e utili, ma diverso è il discorso se si cercano teorie sistematiche e ben organizzate. A parte la terminologia che rischia di spaesare un po' tutti (senso, significato, significazione, effetti di senso, valori semantici, essere “significativo”, eccetera...), bisognerebbe avere più chiaro che siamo di fronte non a un qualche definito tipo di entità ma a un *multiforme insieme* di

(2) Ferraro (2019), pp. 41-42 e 164-168.

processi articolati e anche molto differenziati. Su alcuni di questi abbiamo riflettuto a sufficienza, su altri molto meno. In *Semiotica 3.0*⁽³⁾ avevo constatato la presenza di almeno otto differenti modi in cui un procedimento semiotico può produrre senso — e si tratta probabilmente di un elenco ancora parziale. Non sono differenze di sfumature, ma modalità profondamente diverse, alcune delle quali tra l'altro potrebbero, più facilmente di altre, essere implementate anche nel caso di macchine intelligenti. Si pensi alla produzione di senso basata su connessioni relazionali d'ordine sintagmatico (come vale anche per il vecchio esempio di *nuvole e pioggia*, ma non sempre si tratta di fenomeni così banali!), o al caso, che ci appare sempre un po' misterioso, delle valenze semantiche che hanno origine da puri pattern formali di connessione (per cui eventi slegati non risultano aver senso, ma passano a una condizione di *sensatezza* per il solo fatto di venire formalmente connessi). Mentre per altri casi (meccanismo saussuriano, funzionamento fondativo dei sistemi culturali, eccetera) appare improbabile il coinvolgimento di una macchina — intelligente quanto si vuole, ma dotata di *un altro tipo di intelligenza* rispetto al nostro — i casi citati si avvicinano a modi di effettivo funzionamento dei dispositivi d'IA. E parlare di forme di "senso" in questi termini costituirebbe già un grande passo in avanti rispetto alle vecchie concezioni ferme all'omologazione tra *significato* e *referente*. Di nuovo, questo vale come stimolo per i semiotici, che dei modi di funzionamento dell'aver senso si sono, curiosamente, molto poco occupati — eppure, sembrerebbe doversi trattare di un loro compito istituzionale!

In linea generale (ma anche qui ci sarebbero tanti approfondimenti da fare, da parte dei semiotici) sembra che i dispositivi di IA abbiano difficoltà con le forme discorsive. Possono farvi sopra operazioni quantitative, indicizzarle, tradurle, ma molto più difficilmente comprenderle. Sembrano del resto cavarsela con un solo genere di discorsi, quello meramente informativo. Ammettiamolo, possiamo dichiararci in partenza inferiori a una macchina nell'ardua lettura, poniamo, delle istruzioni di un fornetto multifunzione, ma di fronte a un testo di natura non tecnica e non meramente informativa, emergono chiare le difficoltà che frenano tante forme di software. Non parliamo di testi letterari

(3) Ferraro (2019), pp. 226–229.

o filosofici, ma per constatare tali limiti basta un articolo di giornale, o un testo di storia che vada oltre il mero elenco dei re di Roma. Il punto è sempre quello: le istruzioni di un elettrodomestico riproducono caratteristiche e modi di funzionare di un oggetto reale, si può quindi comprenderle anche con un modello concettuale del significato come semplice rinvio ad un referente. Ma al di là di questo, le operazioni compiute da una mente umana presentano caratteri decisamente diversi. Già in un articolo di giornale, per esempio, il rapporto discorso/realtà assume modalità decisamente più complesse. Tra le altre, vi sono le complicazioni del piano enunciativo, che includono fatti e enunciazioni in un gioco di prospettive e di dispositivi di lettura, di proiezione, di confutazione, e ci sono le complicazioni connesse al gioco delle probabilità, delle ipotesi, delle eventualità e delle previsioni...

Cito in particolare questi aspetti perché si prestano a una considerazione interessante. Da un lato, emergono qui alcune delle ragioni principali per cui gli esseri umani impiegano i sistemi semiotici: vale a dire, per staccarsi dalla realtà attualmente constatabile (si parla non a caso di “pensiero offline”) e introdurre rappresentazioni di stati di cose in qualche modo *virtuali*, ragionando su percorsi di trasformazioni, operando previsioni e confrontando modi di vedere. Esattamente l'opposto rispetto a una concezione referenzialista! Ma in questa direzione scopriamo significative affinità con certe applicazioni di IA, concepite proprio per formulare ipotesi, elaborare scenari virtuali, confrontare stati di cose alternativi, e ovviamente operare previsioni (finanziarie, meteorologiche, di andamento del mercato o di risultati elettorali). Ecco una prospettiva in cui un avvicinamento tra i due ambiti di studi potrebbe risultare molto vantaggioso su entrambi i fronti — paradossalmente, si tratterebbe di spiegare agli informatici che molte volte hanno ragionato su forme di produzione di senso senza rendersene conto! E ancora una volta, ecco uno stimolo significativo per un approfondimento teorico da parte dei semiotici; possibile che ad esempio la teorizzazione del cosiddetto “pensiero offline” sia ora di fatto essenzialmente materia per psicologi o per paleoantropologi? Molto male, perché questo è un terreno in cui la nostra disciplina dovrebbe assumere una posizione decisamente centrale.

Ma siamo dunque al fatto fondamentale per cui, per gli esseri umani, funzioni come “comprendere”, “spiegare”, “assegnare un senso”,

nella vita quotidiana come nella scienza, corrispondono in primo luogo all'operazione che lega tra loro *secondo una relazione definita* determinati elementi, o dati osservativi. Bene, che cosa un software di IA, magari sotto l'etichetta strabiliante di "rete neurale", sa fare meglio di tutto, se non *mettere in relazione*, con la capacità anzi di individuare correlazioni che un essere umano non avrebbe neppure saputo supporre? Anche qui, è evidente la differenza: noi guardiamo immediatamente a correlazioni *significative*, laddove la macchina cerca correlazioni e basta, sensate o meno, comprensibili o meno che possano apparire. Vantaggi e svantaggi del tener conto della dimensione del senso. Perché per la macchina, invece del senso, importano i numeri. Tra l'altro, si parla così spesso di "digitale" da aver dimenticato che "digitale" corrisponde a "numerico" — con le varie e serie complicazioni che ne derivano. La mente umana non è certo aliena dalle competenze matematiche, ma si trova presto a mal partito quando i calcoli minimamente si complicano, sicché nella rappresentazione del reale in cui siamo quotidianamente immersi i numeri, come quelli delle statistiche, hanno un ruolo di fatto limitato: il nostro modo di pensare è fondamentalmente *analogico*, come dimostra anche l'esplosione recente dell'uso di immagini d'ogni genere (e le immagini, ovviamente, sono sempre analogiche; "immagini digitali" possono esistere solo nei peggiori incubi degli informatici).

Nella sua costituzione simbolica e assiologica più profonda, l'universo analogico in cui viviamo (il cui funzionamento è stato illustrato dai classici studi di Claude Lévi-Strauss a proposito di culture prive di scrittura⁽⁴⁾) si regge su una serie di plurimi rinvii di equivalenza (un po' sull'idea della "semiosi illimitata" di Peirce, potremmo dire). "A equivale a B, che equivale a C" è un modo fondamentale di *interpretare* A, e al tempo stesso B e C. I meccanismi della correlazione analogica, misteriosa quanto poliedrica, sfuggono tendenzialmente a una macchina, laddove la correlazione numerica, con l'uso dei famosi *big data*, le consente di legare tra loro certi dati secondo prospettive diverse dalle nostre, ma forse complementari.

Si sta attualmente cercando di rendere le correlazioni digitali significative anche in prospettiva umana: è una direzione interessante, i cui

(4) Per un'approfondita discussione semiotica delle teorie di Lévi-Strauss, v. Ferraro (2001).

risultati al momento ci lasciano però spesso perplessi. Si pensi ad esempio alla traduzione delle tipiche correlazioni statistiche di gradimento in suggerimenti qualitativi: il software può anche non avere la minima idea delle differenze e delle analogie tra due brani musicali di cui ci suggerisce l'affinità ma, percepiti da un orecchio umano, i suoi suggerimenti possono risultare ricchi di senso. Si sa che si può fare, però, molto di più, in particolare attraverso un'opera di etichettatura, i famosi *tags*, sui modi di definizione e gestione dei quali hanno lavorato vari dipartimenti universitari, temo però lasciando del tutto inutilizzata una disciplina, la nostra, che potrebbe forse avere strumenti utili da far valere. È questa in effetti un'altra possibile area di lavoro decisamente interessante per la semiotica, considerando che vi sono aspetti applicativi ma che al tempo stesso ci si colloca anche vicino a grandi temi teorici. In fondo, i modi di classificazione dei testi costituiscono da sempre un tema di sicuro rilievo, a partire dalla classica e mai chiusa discussione sui "generi".

Faccio l'esempio delle "raccomandazioni" di Netflix, avendoci un po' ragionato sopra, insieme ad Antonio Santangelo e Anita Botta (2021). Nel caso di Netflix, appunto, si ha una combinazione tra correlazioni statistiche e *tagging* da parte di soggetti umani. L'analisi da noi condotta ha mostrato come le metodologie attualmente impiegate si fermino a un livello troppo superficiale, che potrebbe essere facilmente superato e arricchito dall'impiego di criteri semiotici anche semplici, come quello della coppia argomentativa *topic/focus*, o di criteri provenienti dalla teoria della narrazione. In questo ambito, che sta assumendo sempre maggior rilevanza, è certamente possibile raggiungere risultati superiori con l'impiego di strumenti di provenienza semiotica — ma, ancora una volta, abbiamo soltanto studi di prima ricognizione.

Legato a questo c'è del resto tutto l'ambito, più generale e articolato, delle tante forme sistematiche di categorizzazione, un tema su cui i teorici dell'IA stanno lavorando in modo decisamente interessante anche dal nostro punto di vista. Penso ad esempio alle sperimentazioni su come le macchine possano definire una categoria a partire da una serie di esempi che non possiedono tutti gli stessi tratti: la macchina costruisce una classe intorno a quello che definisce come "prototipo" del gruppo. In *Semiotica 3.0* io citavo appunto la teoria dei prototipi, sviluppata

peraltro fin dagli anni settanta, come punto di partenza per superare certi aspetti troppo semplicistici del modello saussuriano, pur mantenendone quei caratteri di base che per noi sono fondamentali. Ecco un altro punto su cui il confronto e lo scambio d'idee tra i due ambiti può avere proficui risultati. Anche qui, le logiche da cui si parte saranno ovviamente diverse, muovendo da un lato da concetti cognitivisti, e dall'altro dal nostro — sempre efficace e raffinato, ma in attesa di perfezionamenti — *modello biplanare*, grazie al quale l'identità categoriale degli elementi posti su un piano dipende dalle correlazioni con gli elementi posti sull'altro piano (nei nostri termini, si tratta dei piani del *significante* e del *significato*, o dell'*espressione* e del *contenuto*). Tra l'altro, Luis Prieto (1975) ha mostrato come tale modello valga allo stesso modo nell'ambito del segno e degli utensili, e questo aiuta un avvicinamento con le prospettive funzionali e pragmatiche tipiche dell'IA. Ma sono passati cinquant'anni, e sarebbe una buona cosa che, anche grazie a questa interazione con gli studi sull'IA, ci impegnassimo ad aggiornare questi aspetti assolutamente centrali della teoria semiotica.

5. La dimensione intersoggettiva

Bisogna a questo punto capire perché, tra la visione semiotica e le basi concettuali dell'IA, si ponga un'opposizione di fondo, di natura diremmo filosofica e assiologica. La semiotica, ricordiamolo sempre, è nata come indagine sui meccanismi della traduzione soggettiva del mondo in entità interpretative mentali (Peirce) e come sviluppo del concetto di *soggettività condivisa*, o *collettiva* (Saussure, ma bisognerebbe tener conto anche della semiotica di Emile Durkheim⁽⁵⁾). Oggi, queste radici possono incontrare la teoria della complessità, e cancellare dalla semiotica ogni residuo di un vecchio riduzionismo semplificatore. Pensiamo ad esempio a quella visione della comunicazione che a torto è stata attribuita al povero Roman Jakobson, sintetizzata in quel goffo schema che ha dominato per troppo tempo i manuali di semiotica⁽⁶⁾. Di nuo-

(5) Sulla visione semiotica di Durkheim v. Ferraro (2019), pp. 84-88.

(6) Per una riflessione sul modello di Jakobson e una sua riformulazione attuale, v. Cap. III di Ferraro (2013).

vo, siamo di fronte a un aspetto discutibile e superato della semiotica che, significativamente, presenta notevoli imparentamenti con la teoria cibernetica, madre delle concezioni dell'IA. E sono proprio questi imparentamenti su versanti negativi a poterci aiutare a formulare prospettive più aggiornate e meno ingenuie.

Gli studi sul funzionamento dell'IA tendono a immaginare una macchina che viene posta in quanto tale di fronte a un mondo che deve leggere, analizzare, categorizzare... La semiotica, riprendendo se vogliamo posizioni in senso lato neokantiane, ha da sempre sottolineato che questo rapporto tra il soggetto e il mondo non può essere concepito in modo diretto, né in termini meramente individuali. In particolare, tanto Saussure quanto Durkheim ci hanno dato una nuova concezione della soggettività di natura *istituita e collettiva*. Il salto rispetto alla visione precedente è tanto forte da cancellare il vecchio concetto di "soggettività", facendoci invece pensare che possano essere i dati oggettivi a risultare variabili da caso a caso, da individuo a individuo, mentre le componenti soggettive, che un tempo si tendeva ad escludere in quanto tali dalla conoscenza scientifica, sono quelle condivise, persistenti, sostenute da modelli collettivi: nell'universo socioculturale, è questa dunque la realtà da studiare.

È da escludere l'idea di un rapporto che leghi un soggetto singolo al mondo per il tramite di un elemento simbolico, ad esempio una parola: le entità semiotiche esistono essenzialmente nel contesto di interazioni. Già al livello più semplice, quando una persona parla ad un'altra, bisogna mettere in conto l'azione della cosiddetta "teoria della mente", un concetto che viene dalla psicologia ma che è importante riprendere e tradurre in termini semiotici — dove tra l'altro incontra bene le tante riflessioni che hanno seguito il modo in cui Umberto Eco (1979) ha sottolineato la centralità del concetto di *cooperazione*. Secondo questa prospettiva, è essenziale tener presente il modo in cui ciascun soggetto costruisce *rappresentazioni* delle condizioni interiori della *mente altrui*; si tratta di condizioni relative alle competenze enciclopediche, grammaticali, semantiche, assiologiche eccetera, nonché alle attese e alle intenzioni che presumibilmente caratterizzano coloro con cui entriamo in qualche modo in relazione. Già a questo livello, le strutture di senso non stanno né nella mente di chi parla né in quella di chi ascolta, bensì

in uno spazio d'interazione: non si tratta quasi mai semplicemente di “passare informazioni” ma di provare a *condividere* uno sguardo sulle cose, rendendo accessibile all'uno la percezione del mondo di un altro. Ci rendiamo conto che il senso non è nulla di oggettivo, nulla che stia *là fuori*, e neppure in un *qui dentro* che sia la mente di chi parla e di chi ascolta, bensì in uno spazio d'incontro, una sorta di luogo collocato in un “*tra noi, insieme*”. Scopriamo così che l'*intersoggettività* non è un accessorio concettuale più o meno elegante, ma una base di partenza imprescindibile.

È importante collegare questo anche con i nostri modelli in ambito di teoria della narrazione. Già da alcuni anni sono in corso riconsiderazioni dei ruoli attanziali che sostituiscono le originali definizioni statiche con la dinamica di interazioni complesse⁽⁷⁾, ma è anche vero che se pensiamo al modello dello “schema canonico”, cui si fa tuttora molto spesso riferimento, vediamo ancora un soggetto individuale che autonomamente definisce i propri oggetti di valore, formula il proprio programma narrativo, eccetera — quasi in parallelo, diremmo, al funzionamento del vecchio schema procedurale in ambito informatico: un processo lineare che, a partire dalla definizione degli obiettivi da raggiungere, stabilisce di conseguenza le tappe di un fare realizzativo. Ma né un essere umano né una macchina davvero intelligente agiscono come un soggetto “canonico”.

Nella narrativa di fantascienza — si pensi al caso assolutamente esemplare di *Matrix* — s'immagina che le macchine, piuttosto che banalmente fornire agli esseri umani informazioni, predispongano piuttosto per loro *configurazioni di mondi*, intere *realità* artificialmente costruite. Fanno questo grazie al fatto che le macchine agiscono non come entità separate ma sulla base di un'intelligenza distribuita, e insieme perché vedono il mondo come *fatto di linguaggio* — il loro linguaggio di programmazione innanzi tutto⁽⁸⁾. Questa costruzione narrativa, come si comprende seguendo l'evoluzione della serie di *Matrix*, non fa che proiettare sulle macchine aspetti della società umana che ci è difficile comprendere, o accettare, in forma diretta. Le macchine useranno i loro linguaggi e potranno collegare insieme le loro capacità, cer-

(7) V. ad es. Landowski (2005).

(8) Per una lettura semiotica di *Matrix*, v. Ferraro (2006).

tamente, ma siamo noi umani in primo luogo — dobbiamo intendere — a vivere in un mondo costruito tramite i nostri sistemi semiotici, siamo noi a usare il linguaggio per costruire configurazioni di mondi ed elaborare l'immagine della nostra realtà. E siamo dunque noi umani a muoverci sulla base di un pensiero distribuito, elaborato a livello collettivo. Come tra l'altro si è visto con tutta chiarezza in questi ultimissimi anni, gli esseri umani pensano in forma di aggregazione, per adesione a modelli, siano questi socialmente dominanti o antagonisti. Dovendo scegliere tra quello che appare loro essere più “vero” e quello che invece è realtà sancita a livello condiviso, gli esseri umani tendono più spesso a scegliere la seconda possibilità: perché questo è nella loro natura di animali sociali, sopravvissuti fin dall'origine della loro specie grazie all'abitudine ad agire in branco e a pensare in forme gregarie. Le macchine, invece, non funzionano così. Per quanto si possano fare dei tentativi in senso contrario, come osserva ad esempio anche Justine Lane (2021), le macchine non sono *creature sociali* e non possiedono un'*intelligenza sociale*.

Già: non funzionano così le macchine, e dunque non agiscono in questo modo i software di IA quando impiegati ad esempio nel contesto di una vera e indipendente ricerca scientifica o nel contesto di una seria analisi di dati (economici, sociali, sanitari...): un'analisi che può portare a smentire proprio quelle verità istituite che i membri del gruppo dovrebbero — e spesso vorrebbero — passivamente condividere. Esiste davvero, già oggi, una sorta di posizione eversiva delle macchine, o diciamo meglio un uso alternativo delle risorse dell'IA? Siamo quanto meno di fronte a una vaga contraddizione: un certo impiego dell'IA può di fatto venire a confliggere con l'impiego degli stessi strumenti informatici quando inserito invece nella logica dominante di una socialità fondata sul consenso e sull'aggregazione.

Credo che si disegni oggi — l'ho mostrato recentemente in una ricerca sui modi di percezione del reale durante la pandemia⁽⁹⁾ — uno scontro tra una visione tendenzialmente neoliberista e tecnocratica⁽¹⁰⁾, per la quale il futuro dell'umanità dipende dalla tecnologia, dalle esigenze della macchina economico-finanziaria, da condizioni oggettive

(9) Ferraro (2021a).

(10) Su questo si veda tra l'altro Zuboff (2019).

come quelle attribuite all'ambiente... In tale quadro, fortemente venato da allusioni a una *eterodirezione*, l'IA può essere chiamata a salvarci in certo senso anche da tratti umani visti come troppo arretrati ("l'essere umano è obsoleto"). La visione antagonista contesta tale prospettiva ritenendola una costruzione illusoria, cui contrappone un'oggettività fondata su dati e prove scientifiche, figlia di quel modello comunicativo a partecipazione orizzontale che sarebbe promosso dalla rete. Come si vede, un'ideale *oggettività* può essere declinata in chiavi opposte, e la cosa interessante per noi, in questa sede, è che correlazioni statistiche, analisi di dati automatizzate, e la stessa "indipendenza di pensiero", diciamola così, attribuibile alle macchine, possono venire coniugate con la dimensione sociale in modi radicalmente differenti.

Ma attenzione, la posizione "tecnocratica" è oggi invisibile anche ad altri, proprio per il suo pericoloso e inaccettabile *oggettivismo*. Si tratta in questo caso di una prospettiva decisamente diversa, basata sull'ipotesi di un rinnovamento socioculturale che di fatto si richiama alla dimensione del senso, sottolineando la necessità di innovare i nostri codici di lettura del reale a partire dai valori in cui crediamo e da una riflessione sui processi interpretativi tramite i quali definiamo il nostro rapporto con ogni aspetto del mondo che ci circonda. A una prospettiva di eterodirezione, si contrappone l'idea di un soggetto collettivo consapevole delle proprie scelte.

Questa contrapposizione ci riconduce alla questione della differenza tra uomo e macchina, e dell'insuperabile eterogeneità di fondo tra la nostra intelligenza e quella artificiale — temi che costituiscono il forse quasi inevitabile punto d'arrivo di un percorso di riflessione semiotica. Perché possiamo dare a una macchina la capacità di parlare, di dipingere e di fare musica, di raccontare storie e così via, ma queste varranno sempre come specifiche componenti aggiuntive: non, com'è invece per noi, quali nostri tratti essenziali. Alla fine, che cosa definisce un essere umano? Ed è questo un tema, davvero, fondamentalmente semiotico?

Ho avuto per parecchi anni l'abitudine di affrontare tale argomento nella lezione inaugurale del corso di Semiotica⁽¹¹⁾, avvalendomi tra l'altro del confronto oppositivo proposto da Herbert George Wells (1898) grazie al disegno, diventato poi archetipico, dell'alieno-non-umano:

(11) V. Ferraro (2013), pp. 35-37.

un essere tutto razionalità, e in questo senso puro e perfetto, gelido cervello scevro da emozioni e desideri, che non necessita di inghiottire cibo né di avere alcun contatto con i suoi simili o con un corpo materno, e che comunica solo per via telepatica: dunque un essere che non ha esperienza di alcun rapporto significativo con una dimensione materiale. In questo ritratto in negativo c'è un'implicita ma magistrale definizione di ciò che noi siamo, magistrale perché mette in collegamento tante cose: il nostro rapporto con le materie di cui ci cibiamo con l'uso di segni a manifestazione sensibile, la relazione primaria con il corpo materno con la tendenza a provare emozioni, la nostra differente razionalità con l'attitudine a sognare e ad illuderci, il tutto logicamente connesso in un disegno d'insieme ove ogni aspetto s'intreccia indissolubilmente con gli altri. Mi piace inserire anche, in questo disegno, la nostra capacità di creare magiche opere d'ingegno tramite un rapporto fisico con la materia, ad esempio sporcandoci le mani con le paste colorate che stendiamo su una tela, o magari producendo suoni grazie allo sfregamento di crini di cavallo su budelli animali (così, almeno in origine, funzionavano gli strumenti ad arco). È questo connubio inscindibile e continuo di intelligenza e corporeità, di immateriale trascendenza di senso e di indispensabile, ruvido contatto con gli oggetti tangibili, a renderci esseri umani, intrisi di una primaria, irrefrenabile propensione alla semiosi.

Per quanto io non intenda proporre una piena analogia tra i gelidi alieni di Wells e i nostri automi a intelligenza artificiale, credo che la contrapposizione sia chiara, e la chiave della differenza inevitabilmente forte: una macchina intelligente non condividerà mai l'esperienza che ci rende umani, né condividerà mai quell'irriducibile legame che per noi fonde insieme il mondo sensibile e l'universo del senso. Non si tratta, intendiamoci, di uno scontro tra *bene* e *male*. La nostra specie, caratterizzata da questo rapporto immediatamente interpretativo con il reale, può certo essere vista come insieme meravigliosa e detestabile, dotata di grandi forze creatrici e di non minori tendenze distruttive, ma se vogliamo valorizzarne i lati più positivi, la cosa meno utile da fare è rischiare confusioni, o forzare le analogie tra quelle che, dobbiamo riconoscerlo, non sono soltanto *forme diverse d'intelligenza*. Sappiamo del resto che l'espressione "intelligenza artificiale" è giudicata da molti poco chiara. Se cerchiamo di comprendere il valore dei termini secondo il principio saussuriano

dell'identità negativo–differenziale, possiamo dire che “artificiale” si oppone non al termine “naturale” (di significato ampiamente discusso) ma a “biologico”, e “intelligenza” dal canto suo si differenzia da “organismo”. Il confronto tra i concetti di “intelligenza artificiale” e “organismo biologico” dice molte cose. Non solo che l'una è progettata dall'uomo e l'altra no, ma che l'una costituirà sempre una *parte*, artificialmente e un po' artificiosamente circoscritta, mentre l'altra possiede i caratteri di una totalità strutturale ove tutte le componenti sono interconnesse e interdefinite. Come osserva Éric Sadin (2018: 12 e 21), architetture computazionali prive di corpo, per quanto collegate a sensori di vario tipo, restano comunque escluse da un'infinità di dimensioni che la nostra sensibilità è in grado di cogliere; nulla risulta più illusorio di un “antropomorfismo frammentario”. In effetti, che si possa prendere una porzione, l'intelligenza, considerandola a parte, è un puro *mito*, e un mito che significativamente ci ripresenta un po' i caratteri di quel funzionalismo che nella prima metà del novecento aveva imperato, nel design come nell'architettura (pensiamo a tratti come la semplificazione estrema, la cancellazione dell'inessenziale, la tendenza alla razionalità pura e all'economicità concettuale, con riduzione del disegno globale allo scopo definito). Gli esseri umani, che uno scopo definito appunto non ce l'hanno, non possono condividere questi tratti, e non possono affidare alle macchine il disegno *globale* delle cose. Possono fare uso, invece, delle macchine, proprio perché queste sono dotate di capacità *limitate e definite*, volutamente *settoriali* e progettualmente *semplificate*.

Riferimenti bibliografici

- ECO U. (1979), *Lector in fabula. La cooperazione interpretativa nei testi narrativi*, Bompiani, Milano.
- FERRARO G. (2001), *Il linguaggio del mito. Valori simbolici e realtà sociale nelle mitologie primitive*, nuova ed. Meltemi, Roma.
- (2006), “Matrix: l'anomalia e il sistema”, in N. Dusi (a cura di), *Remix–Remake*, Meltemi, Roma, 279–300.
- (2013), *Fondamenti di teoria sociosemiotica. La visione “neoclassica”*, Aracne, Roma .

- (2019), *Semiotica 3.0*, Aracne, Roma.
- (2021a), “L’incidente e il sistema. Forme di narrazione dell’epidemia”, in *Acta Semiotica*, 1: 104–125.
- (2021b), “Modèles classiques et complexité sémiotique”, in *Acta Semiotica*, 2: 29–40.
- (2022), “Le rythme comme règle et comme invention”, in *Acta Semiotica*, 3: 21–44.
- FERRARO G., Santangelo A., Botta A. (2021), “Il significato di un suggerimento di visione. Riflessioni semiotiche sul sistema di raccomandazione di Netflix”, in *DigitCult* 37, Vol. 6, 1: 37–50.
- GREIMAS A.J. (1976), *Maupassant. La sémiotique du texte: exercices pratiques*, Seuil, Paris (trad. it. *Maupassant. Esercizi di semiotica del testo*, Bompiani, Milano 2019).
- KYROU A. (2020), *Dans les imaginaires du futur: entre fins du monde, IA, virus et exploration spatiale*, Éditions ActuSFm, Paris.
- LANDOWSKI, E. (2005), *Les interactions risquées*, Pulim, Limoges.
- LANE J.E. (2021), *Understanding Religion through Artificial Intelligence*, Bloomsbury Academic, London.
- PRIETO L. (1975), *Pertinence et pratique*, Minuit, Paris (trad. it. *Pertinenza e pratica*, Feltrinelli, Milano 1976).
- SADIN É. (2018), *L’Intelligence artificielle ou l’enjeu du siècle. Anatomie d’un antihumanisme radical*, L’échappée, Paris (trad. it. *Critica della ragione artificiale. Una difesa dell’umanità*, Luiss University Press, Roma 2019).
- TOMLINSON G. (2015), *A Million Years of Music. The Emergence of Human Modernity*, Zone Books, New York.
- WELLS H.G. (1898), *The War of the Worlds*, Heinemann, Portsmouth (trad. it. *La guerra dei mondi*, Mursia, Milano 2009).
- ZUBOFF S. (2019), *The Age of Surveillance Capitalism*, PublicAffairs, New York (trad. it. *Il capitalismo della sorveglianza*, Luiss University Press, Roma 2019).

IL MITO DELLE DUE INTELLIGENZE E LA SEMIOTICA

UGO VOLLI*

ENGLISH TITLE: *The Myth of the Two Intelligences and Semiotics*

ABSTRACT: The paper aims to consider the contrast between human and artificial intelligence. Lexicographic analysis shows that intelligence is a process or activity destined for adaptation by necessarily imperfect human individuals. This relationship between intelligence and imperfection, confirmed by experience, is however unsatisfactory from a philosophical and theological point of view. Hence a long series of attempts to establish exact mechanisms of artificial intelligence, capable of automatically formalizing semantics and drawing irrefutable truths from them. What we call artificial intelligence, however, is not the realization of these attempts, but more modestly the sense effect of the automatic classification of large data sets, based on human examples, which allow software feedback. These are mechanisms of great interest for semiotics as well as for cognitive sciences and neurosciences.

KEYWORDS: Artificial Intelligence, Semiotics, Intelligence, *Machine learning*, Leibniz

I nomi contano. Nella nostra società da una decina d'anni è emerso di nuovo con forza quello dell'"intelligenza artificiale", descritta come scientifica, esatta, pressoché infallibile, e sostanzialmente migliore di quella umana. Di "intelligenza naturale" in realtà in questo periodo si parla poco, almeno sul fronte della cultura generale, se non su quello specialistico della ricerca psicologica, anche perché gli ultimi dibattiti pubblici sul tema, una ventina d'anni fa, sono stati spesso bloccati per

* Università degli Studi di Torino.

ragioni di “correttezza politica”, quando dalle prove sperimentali emergeva che essa non è certo uguale per tutti gli esseri umani. La mutazione della psicologia nelle “scienza cognitive” ha lasciato in ombra questo tema e anche più radicalmente la *vexata questio* della definizione di “intelligenza”, in particolare della sua struttura unitaria o modulare che era stata centrale negli ultimi decenni del secolo scorso (Fodor 1983, 2000) con riflessi anche in semiotica, soprattutto in Eco e nella sua scuola. Questi sono però temi che a mio avviso esulano dall’ambito disciplinare della semiotica e non li affronterò in questo articolo. Ciò che mi interessa è capire come certe funzioni delle macchine elettroniche siano state definite “intelligenza”, creando così la duplicità fra uomo e macchina in questo ambito. Comunque si definisca l’intelligenza umana, si tratta chiaramente di un’estensione metaforica, ma proprio questa dimensione retorica interessa la semiotica e vale la pena di essere interrogata.

Nell’affrontare questo tema, può essere utile tener presente un libro che è stato popolarissimo e molto influente negli anni ‘60 e oggi non si conosce quasi più, ma a me pare ancora attuale, *Le due culture* (*The Two Cultures and the Scientific Revolution*), del fisico e intellettuale britannico Charles Snow, pubblicato in inglese nel 1959 e subito tradotto in italiano. Scrive Snow (1959: 14):

sono convinto che la vita intellettuale, nella società occidentale, si va sempre più spaccando in due gruppi contrapposti [...]: a un polo abbiamo i letterati, che come per caso, senza che nessuno se ne accorgesse, cominciarono ad autodefinirsi “intellettuali,” quasi che non ce ne fossero altri [...] scienziati all’altro, i più rappresentativi dei quali sono i fisici. Tra i due gruppi, un abisso di reciproca incomprensione [...] I non-scienziati hanno una radicata impressione che gli scienziati siano animati da un ottimismo superficiale e non abbiano coscienza della condizione dell’uomo. D’altra parte, gli scienziati credono che i letterati siano totalmente privi di preveggenza e nutrano un particolare disinteresse per gli uomini loro fratelli; che in fondo siano anti-intellettuali e si preoccupino di restringere tanto l’arte quanto il pensiero al momento esistenziale.

Ancora oggi senza dubbio ci sono due culture che spesso si ignorano e si disprezzano, anche se la prevalenza di un modello matematizzante

nelle scienze umane è evidente; ma, come ho detto, sembrerebbe che ci siano anche due forme di intelligenza: una “naturale” e “umanistica” e una “artificiale” e “tecnologica”. È una contrapposizione senza dubbio in continuità con quella fra le due culture.

Tenendo presente dunque la dimensione polemica sottolineata da Snow e ancora ben viva, vale la pena di cercare di comprendere non che cosa *sia* l'intelligenza — un tema scientifico particolarmente complesso e certamente non affrontabile con gli strumenti della semiotica —, ma che cosa *si intenda* nella nostra cultura (non diversamente da altre di origine europea), con questa parola; vale a dire quale *unità culturale* (Eco 1984: 178-179) venga mobilitata utilizzando il termine “intelligenza”. Dal punto di vista etimologico è chiaro che la parola italiana, come i calchi più o meno perfetti che troviamo in inglese, francese, spagnolo, portoghese, tedesco, russo e perfino in ebraico moderno, deriva, come altre voci lessicali fra cui *intellectus*, dal verbo latino *intelligo* o *intellĕgo*, che a sua volta viene da *inter* e *lego*, cioè “collego assieme”.

Vale la pena di leggere qualche definizione di questo termine.

Sabatini Colletti: intelletto, ingegno; prontezza e vivacità di mente; abilità e perizia nel fare qualcosa.

De Mauro: facoltà della mente umana di intendere, pensare, giudicare, comunicare fatti e conoscenze, di formulare giudizi ed elaborare soluzioni in risposta agli stimoli esterni, di adattarsi all'ambiente o di modificarlo in base alle proprie necessità

Treccani: complesso di facoltà psichiche e mentali che consentono di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e adattarsi all'ambiente.

Tommaseo: facoltà intellettuale, abito e atto e modo dell'intendere

Garzantina Filosofia: in un'accezione ampia, processo mentale che consente all'uomo o all'animale dotato di struttura celebrale evoluta la soluzione di problemi nuovi che implicano una ristrutturazione del rapporto adattativo all'ambiente; in un'accezione specifica, processo o insieme di processi mentali molto complessi e specificamente umani, dal ragionamento logico alla capacità di perseguire uno scopo anche a lunghissimo termine, alla capacità di formulare valutazioni e giudizi di valore, alla capacità di autocorrezione e autocritica.

Se consideriamo il versante del significato di questa unità culturale come si presenta nelle diverse lingue, troviamo una costanza analoga a quella del significante. Per esempio in inglese, “the ability to learn or understand or to deal with new or trying situations; the skilled use of reason; the ability to apply knowledge to manipulate one’s environment or to think abstractly as measured by objective criteria (such as tests)”⁽¹⁾. E in francese, “ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle; aptitude d’un être humain à s’adapter à une situation, à choisir des moyens d’action en fonction des circonstances”⁽²⁾.

Emergono qui tre aspetti che mi sembrano significativi: il radicamento nella dimensione *umana* o al massimo animale dell’intelligenza; il suo essere *non un oggetto*, per quanto astratto (non uno stato o una qualità o una cosa), bensì una *facoltà* o un *processo*, cioè un fenomeno dinamico; e infine il fatto che esso è *in vista di qualcosa*, ha una dimensione teleologica.

Ciò a cui tale facoltà è diretta è ovviamente la *conoscenza*, ma più spesso in generale *l’adattamento* al mondo di chi ne è portatore, quel che in termini darwiniani⁽³⁾ è la sua *fitness*. Essendo umana e relativa a scopi, l’intelligenza è *preziosa* ma *sogettiva* e *prospettica*, dunque *fallibile* e spesso *debole*.

Mentre filosofia e teologia parlano spesso di *intelletto attivo* nella divinità (νοῦς ποιητικός) o *saggezza divina* (חוכמה, Σοφία), l’intelligenza, nelle sue varie versioni (φρονεσις, μέτις, ללאקסימ) non ha dunque affatto carattere assoluto o divino, proprio per il suo carattere di *mezzo soggettivo* diretto a un *fine*, che dunque presuppone una *mancaza*, un *desiderio* o almeno un *bisogno* (Volli 2002). Questo carattere imperfetto, *tentative*, approssimativo dell’intelligenza, la caratterizza come una di quelle doti, il cui carattere “intermedio” fra divinità e animalità (*metaxy*), da Platone (*Simposio* 202e–203a; *Timeo* 50d) a Pico della Mirandola (*De hominis dignitate*) alla filosofia contemporanea, caratterizza la definizione dell’umano.

(1) www.merriam-webster.com/dictionary/intelligence (ultima consultazione 23 agosto 2022).

(2) www.larousse.fr/dictionnaires/francais/intelligence/43555 (ultima consultazione 23 agosto 2022).

(3) Ma in realtà il termine è di Herbert Spencer, che lo usò per spiegare il metodo usato da Darwin nell’*Origine delle specie*; essa fu successivamente accettata da Darwin.

Ma naturalmente l'imperfezione è insoddisfacente: per questo è coestensiva al desiderio. Chiunque sia critico dello stato del mondo, implicitamente tende a considerare insufficiente l'intelligenza che lo regge e gli dà forma: dei governanti, del popolo, dei nemici, la sua stessa; ma anche quella della scienza, che è per definizione imperfetta e approssimativa, e spesso perfino quella di Dio che si esprime nel mondo. Come spiega Kant, "per teodicea si intende la difesa della saggezza suprema del Creatore contro le accuse che le muove la ragione a partire dalla considerazione di quanto nel cosmo vi è di contrario al fine di questa saggezza"⁽⁴⁾.

Questa domanda sull'adeguatezza dei pensieri, dei comportamenti, delle istituzioni è uno dei temi principali della riflessione filosofica, il punto di contatto fra la sua *pretesa di conoscenza* sulla struttura della realtà e l'altra pretesa, diffusa a partire dalle origini presocratiche, a insegnare agli esseri umani a *migliorarsi sulla via della saggezza* e a indicare le migliori organizzazioni sociali. Questo tema viene sviluppato in molti modi, per esempio come insegnamento della saggezza personale, come critica del linguaggio, come teoria politica e della giustizia. Ma quel che interessa noi ora è l'idea di coltivare un *miglioramento dell'intelligenza*, da realizzarsi usando vari formalismi o meccanismi. Si tratta della tentazione permanente (o se si vuole del progetto utopistico) di emendare l'intelligenza dai suoi limiti umani, per renderla *logicamente perfetta*, magari *matematica*, soprattutto *oggettiva e automatica*. L'intelligenza completa non può e non deve essere *umana*, affetta cioè dall'*imperfezione*, dal carattere "intermedio", dalla soggettività, dalla subordinazione a fini concreti di adattamento che caratterizzano la condizione umana, ma non può neppure essere *divina*, perché non è possibile convocare la divinità per ottenerne la comprensione; dunque dev'essere *artificiale*.

È un mito semiotico che ricorda quello delle lingue perfette, su cui ha lavorato a lungo Umberto Eco (1993) e la *misologia* dei filosofi (Volli 1993) che ne deriva inevitabilmente. La teoria logica di Aristotele e dei suoi successori, fino alla scolastica, col suo tentativo di normare su basi puramente formali il ragionamento valido, di modo che da premesse certe si arrivi a deduzioni corrette di conseguenze altrettanto vere, va intesa anche in questo senso.

(4) Kant, I. (1994 [1791]), "Sull'insuccesso di ogni saggio filosofico di teodicea", in Id., *Scritti di filosofia della religione*, a cura di Giuseppe Riconda, Mursia, Milano, p. 53.

Possiamo però individuare la prima tappa compiuta di questo progetto (o mito) nell'*Ars compendiosa inveniendi veritatem seu ars magna et maior* di Ramon Lull (1274). Lo scopo di Lull era quello di costruire una tecnica automatica e materializzata del pensiero, un vero e proprio *dispositivo pensante* che, servendosi anche di schemi e figure, potesse collegare concetti fondamentali in una sorta di *logica meccanica*, in modo da ottenere verità in ogni campo del sapere. Non sono le *conseguenze* valide l'obiettivo di questo strumento, non si tratta di una materializzazione degli *schemi sillogistici* già catalogati esaustivamente dalla Scolastica, ma proprio di un sistema per ricavare automaticamente in maniera valida le *premesse* vere di ogni ragionamento, che secondo Lull si dovrebbero ottenere da uno studio dei concetti primitivi e dalla loro combinazione per via di permutazione. In questa maniera egli coltiva un progetto estremamente ambizioso sul piano teorico, oltre che su quello pratico: dimostrare incontestabilmente le verità del cristianesimo anche di fronte a “infedeli” come ebrei e musulmani, allo scopo di ottenerne la conversione. Il progetto di evangelizzazione per via discorsiva delle altre grandi religioni monoteiste non è nuovo, ha un grande momento esemplare mezzo secolo prima (1219) col viaggio di San Francesco in Egitto; e neppure è innovativo il confronto dialettico fra le religioni monoteistiche, oggetto di un libro importante come il *Sēfer Ha-Kūzārī* di Yehuda Ha-Levi, scritto intorno al 1140⁽⁵⁾. Come spiega Paolo Rossi,

nella tradizione del lullismo c'è l'idea di arrivare a una enciclopedia totale, a un sapere universale. Cioè, ho una “clavis” universale, cioè una “chiave” universale, che mi permette di accedere a qualunque sapere nella sua totalità. Quindi dietro ai grandi teatri del mondo del Cinquecento, per esempio in Gimma e tanti altri, che erano, addirittura, delle costruzioni fisiche, cioè erano degli oggetti di legno, che si potevano far ruotare variamente e potevano dar luogo a tutte le combinazioni possibili, c'è questo sogno strano — ma gli uomini perseguono sogni strani — di arrivare a un mezzo che possa dare accesso alla totalità del sapere. È un sapere tutto organico. Ho tutti gli elementi e rimettendoli insieme posso ricostruire tutto in tutte le sue possibili varietà e in tutte le sue possibilità.⁽⁶⁾

(5) Per una discussione sul rapporto fra Lull e Ha-Levi, rimando a Volli, U. e V. Robiati Bendaud (2021).

(6) <https://web.archive.org/web/20170406201351/http://www.emsf.rai.it/aforismi/aforismi.asp?d=316#links> (ultima consultazione 23 agosto 2022).

Senza entrare nei dettagli, possiamo considerare eredi di questa intuizione Charles de Bovelles (1475-1566), che propone con la sua *Ars operatorum* una tecnica automatica per scoprire tutte le verità, e in seguito le mnemotecniche rinascimentali (per esempio quella di Giordano Bruno) brillantemente studiate da Francis Yates (1966), che segnano un significativo, benché provvisorio slittamento dei dispositivi automatici di associazione concettuale dall'ambito dell'intelligenza a quello della memoria.

Ma soprattutto è rilevante il progetto di *Characteristica universalis* che Leibniz probabilmente trasse da queste fonti e che cercò di sviluppare durante tutta la sua vita, iniziando fin dal lavoro giovanile *De arte combinatoria* (1666). Questi tentativi (anche in Leibniz) hanno un obiettivo argomentativo, politico e teologico, quindi retorico, ma vogliono raggiungerlo realizzando una *logica dei contenuti interamente calcolabile*. Ricordo qui una frase molto celebre del filosofo tedesco: “non ci sarà più bisogno fra due filosofi di discussioni più lunghe di quelle tra due matematici, poiché basterà che essi prendano le loro penne, che si siedano al loro tavolo (riferendosi, se lo desiderano, a un amico) e che entrambi dicano: ‘Calcoliamo!’”⁽⁷⁾.

Il punto qui è la *sostituzione del pensiero discorsivo col calcolo*. Come già aveva sostenuto Hobbes, il pensiero consiste nella manipolazione di simboli, il che implica che esso è spiegabile con leggi scientifiche, se solo si riesce a individuarne le leggi. Scrive ancora Leibniz:

le lingue ordinarie, sebbene siano assai utili al ragionamento, sono tuttavia soggette a innumerevoli equivoci e non possono sostituire il calcolo, in modo cioè che gli errori di ragionamento possano essere scoperti dalla stessa formazione e costruzione delle parole [...] A me invero, mentre meditavo più profondamente su questo argomento, apparve manifesto che tutti i pensieri umani potevano risolversi completamente in pochi pensieri da ritenersi come primitivi. E che se si assegnano a questi ultimi dei caratteri, da essi si possono formare i caratteri delle nozioni derivate, dai quali sarà sempre possibile ricavare i loro requisiti e le nozioni primitive in essi racchiuse [...] Ora una volta che

(7) Leibniz, G.W. (1688) “De arte characteristica ad perficiendas scientias ratione nites”, in *Sämtliche Schriften und Briefe*, a cura della Accademia delle Scienze di Berlino, Serie VI, *Philosophische Schriften*, 1677-Juni 1690, vol. 4, pp. 912-13.

ciò fosse realizzato, chiunque nel ragionamento e nello scrivere facesse uso di siffatti caratteri o non sbaglierebbe mai oppure riconoscerebbe da sé i propri errori non meno che quelli degli altri mediante esami facilissimi.⁽⁸⁾

Il progetto di Leibniz in effetti si rivelò irrealizzabile, come era accaduto per quello di Lull. Il punto non è solo tecnico, non riguarda la difficoltà di realizzare dei meccanismi complessi come quelli richiesti dalla manipolazione fisica dei “caratteri”, che sarebbe stata superata solo dalla meccanica di precisione ottocentesca e soprattutto dall’elettronica nel Novecento. Vi è qui un *problema semiotico*. Il progetto di Leibniz (e di Lull e di Hobbes) ha come punto cruciale la rappresentazione univoca dei “pensieri” con “caratteri” (o “simboli” nel linguaggio di Hobbes) cioè l’assegnazione di un sistema *non ambiguo ed esaustivo* di significanti ai “pensieri” (cioè ai significati), togliendo ad essi la loro “equivocità”, stabilendo fra loro delle relazioni primitive e rendendoli così “calcolabili”. Ora, Eco (1984: 91-106) ci ha insegnato che non è possibile un’analisi esaustiva e inequivoca del piano del contenuto, che l’albero di Porfirio non può classificare i significati senza cadere in contraddizioni e strani anelli (Hofstadter 2007), ma quest’analisi impossibile è esattamente il cuore del progetto leibniziano/lulliano, con in più il requisito fortissimo della *calcolabilità* effettiva.

Leibniz si impegnò direttamente a costruire una macchina per questo calcolo, ma la difficoltà materiale della costruzione di un dispositivo del genere erano molto al di là delle capacità ingegneristiche e metallurgiche del Seicento; i primi progressi effettivi sul piano tecnico furono fatti dopo un paio di secoli con Ada Lovelace e Charles Babbage, che però molto opportunamente limitarono il loro progetto al calcolo numerico.

Anche la logica nel periodo moderno si matematizza e perde definitivamente ogni ambizione di contenuto (salvo in teorizzazioni molto particolari come la logica del senso di Deleuze, che non ha mai ambito alla calcolabilità). Come il progetto delle lingue perfette, anche quello del calcolo del pensiero sembra insomma necessariamente utopico, perché cerca di sovrapporre due cose incompatibili: il funzionamento della semantica umana, che è locale, approssimativo, finalizzato, affetto da

(8) Leibniz, G.W. (1968) *Scritti di logica*, Zanichelli, Bologna, p. 241.

circolarità ma capace di gestirla, produttivamente equivoco, con il formalismo del calcolo: universale, necessariamente non contraddittorio, inequivoco per definizione.

La pretesa che un meccanismo logico–matematico potesse esprimere intelligenza fu riaperta solo a metà del ‘900 con Alan Turing. L’idea però non era più retorico–metafisica, ma *imitativa/simulativa*. Si trattava di costruire una imitazione/simulazione (dunque un’*icona*) dell’intelligenza umana, non una sua sostituzione. Turing dedica all’“Intelligent Machinery” un articolo del 1948⁽⁹⁾. Prima di esaminarlo brevemente, vale la pena di precisare che l’espressione “Artificial Intelligence” compare per la prima volta in una call formulata da John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon nel 1955:

we propose that a 2–month, 10–man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.⁽¹⁰⁾

Il termine chiave, dal nostro punto di vista, è qui *simulare*. L’intelligenza artificiale si presenta come una simulazione sufficientemente precisa del comportamento degli umani, capace perciò di descriverla adeguatamente (*tesi dell’Intelligenza Artificiale debole*) o di eguagliarla (*tesi forte*). Tutto ciò nasce dall’idea estremamente brillante di Turing di proporre un “test” linguistico per l’Intelligenza Artificiale. Vale la pena di leggere qualche riga di questo articolo fondamentale:

I propose to consider the question, “Can machines think?” This should begin with definitions of the meaning of the terms “machine” and

(9) <https://weightagnostic.github.io/papers/turing1948.pdf> (ultima consultazione 23 agosto 2022).

(10) McCarthy, J., Minsky, M., Rochester, N., Shannon, C.E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf> (ultima consultazione 23 agosto 2022).

“think”. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. [...] Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game which we call the “imitation game”. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. [...] We now ask the question, “What will happen when a machine takes the part of A in this game?” Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, “Can machines think?”⁽¹¹⁾

Non entro qui nei dettagli analizzati da Turing, né nel dibattito successivo, al cui centro vi è un famoso controesempio di Searle, quello della “stanza cinese”⁽¹²⁾. Quel che bisogna sottolineare però è che la simulazione entra in due aspetti o livelli, come test di adeguatezza e come progetto sugli obiettivi delle nuove macchine o programmi. Bisogna sottolineare che il progetto di Turing non è pratico/utilitario, bensì teorico: si tratta di capire se l’intelligenza artificiale sia possibile e soprattutto come riconoscerla in maniera sufficientemente chiara e oggettiva. Non si tratta di convincere nessuno, ma di fare cose che solo gli umani sanno fare, magari superandoli. Per questo già Turing capisce che è essenziale la capacità di imparare, modificando se stessi: “What we want is a machine that can learn from experience,” and that the “possibility of letting the machine alter its own instructions provides the mechanism for this.”⁽¹³⁾

L’intelligenza artificiale è dunque il mito, il punto di fuga. Ma l’aspetto determinante è il *machine learning*, la capacità di apprendimento

(11) *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460, <https://doi.org/10.1093/mind/LIX.236.433> (ultima consultazione 23 agosto 2022).

(12) Searle, J.R. (1980) *Minds, brains, and programs. Behavioral and Brain Sciences*, Volume 3 (Issue 3): pp. 417–457, <https://web-archival.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf> (ultima consultazione 23 agosto 2022).

(13) <https://www.britannica.com/technology/artificial-intelligence/Evolutionary-computing> (ultima consultazione 23 agosto 2022).

di un non-soggetto, che produce effetti soggettivi. Questa dimensione diventa sempre più importante quando il progetto si distacca dall'ambito teorico per assumere finalità pratiche. Il *machine learning* in sostanza è un processo di inferenza capace di retroazione sui parametri. In termini echiani/peirceani possiamo considerarlo come un'implementazione del meccanismo di abduzione che sta alla base del funzionamento semiotico-cognitivo. Se la semiotica si impegnasse a seguire questa connessione probabilmente potrebbe contribuire in maniera significativa a questa esplorazione sull' "intelligenza".

Il meccanismo inferenziale retroattivo del *machine learning* interessa molto noi semiotici, perché genera un *effetto* analogo al funzionamento testuale. Molte volte, dato un certo universo di oggetti (risultati, immagini, conversazioni) — nei nostri termini: parti di un testo —, all'Intelligenza Artificiale si assegna il compito di *classificarli* efficacemente, per lo più prendendo come base o esempio analoghe attività umane. Si tratta di un problema di concettualizzazione o categorizzazione, di quelli che interessavano moltissimo a Eco (si pensi al Modello Q del *Trattato*, ai problemi di *ratio facilis* e *difficilis*, ecc.).

Dunque al di fuori della mitologia dell'"intelligenza" sostantiva, dobbiamo riconoscere che le varie "IA" sono programmi costruiti per procedere per tentativi e errori a classificare i dati e interagire con essi, potendo modificare certi parametri cruciali. Vi sono diversi modi per ottenere questi effetti, per esempio le reti neurali, i programmi che misurano la distanza dei dati da un certo parametro cercando di ottimizzarla, ecc. Tutto ciò è di grande interesse, ma è soprattutto importante capire che non vi è alcun magico ingrediente di "intelligenza" che si intrometta in questi risultati, solo un design dell'algoritmo basato sul *feed back* e una certa strategia di classificazione e ottimizzazione.

L'altro aspetto molto significativo è che in queste versioni più recenti e simulate, la mimesi dell'intelligenza viene prodotta rispetto agli *effetti*, cioè ai testi e ai comportamenti visibili: non vi è affatto pretesa di riprodurre la *semantica* dell'intelligenza umana, di trovare dei significanti inequivoci per i termini primitivi del pensiero: un modo di procedere lulliano o leibniziano che ancora qualche decennio fa ispirava l'idea del "mentalese", l'autentico linguaggio della mente che secondo Fodor (1994) risiedeva da qualche parte nella nostra mente e

motivava la nostra comprensione del linguaggio e la nostra intelligenza. L'intelligenza artificiale, se vogliamo ancora chiamarla così, non è più una "macchina per pensare" in stile leibniziano o lulliano e neppure una simulazione dell'intelligenza, in stile turinghiano, è un *insieme di tecnologie* che permettono ai calcolatori di *riconoscere pattern* e configurazioni sulla base di *esempi* offerti durante un "allenamento" per prova ed errori. Il risultato ci appare "intelligente" perché offre una categorizzazione di dati spesso molto numerosi e sfuggenti, secondo criteri *non stabiliti esplicitamente in precedenza*. Ma si tratta essenzialmente di statistica (Bishop 2006, Alpaydin 2020, Vapnik 1998).

In questo *funzionamento non semantico* dell'Intelligenza Artificiale contemporanea, un semiologo può scorgere dei suggerimenti molto interessanti. L'Intelligenza Artificiale, per brillanti che siano i suoi risultati, appare insomma come un *effetto di senso*, non come la costruzione di una mente caratterizzata dalla manipolazione di idee chiare e distinte. Bisogna aggiungere però che, se guardiamo le cose freddamente, *anche* l'"intelligenza naturale" è probabilmente un *effetto di senso*. Oggi i neuroscienziati stanno incominciando a comprendere almeno in parte le strategie adattative che portano alle capacità "intelligenti" di categorizzazione e interazione degli esseri viventi e degli umani. Ridotte alla radice, anch'esse sono in sostanza dispositivi che producono senso per via induttiva, massimizzando certi parametri. Insomma, oggi è chiaro che neanche nel nostro cervello c'è un fantasma nella macchina (Gilbert Ryle) che scruterebbe il mondo con intelligenza trascendente.

Ma se guardiamo ai modi in cui la semiotica considera il senso, all'idea di configurazioni di *profondità* (come il percorso generativo) o di *latenzialità* (come la deriva degli interpretanti) che sono indispensabili al funzionamento dei meccanismi semantici, anzi li costituiscono, vediamo che anche dal punto di vista semiotico il requisito della chiarezza e distinzione dell'appercezione (com'è delineata esplicitamente in filosofia da Cartesio a Husserl, ma con radici che risalgono fino alla metafora della "contemplazione" del mondo delle idee in Platone) è un *mito consolatorio*. Proprio perché è in grado di riconoscere e analizzare la differenza fra effetti di senso e meccanismi di produzione semiotica, proprio perché è abituata a guardare anche al pensiero come montaggio di segni o testi, la semiotica può essere utile per comprendere l'intelligenza come un effetto di senso.

Riferimenti bibliografici

- ALPAYDIN E. (2020), *Introduction to Machine Learning*, MIT Press, Boston.
- BISHOP, C. (2006) *Pattern Recognition and Machine Learning*, Springer, Berlin
- ECO, U. (1984), *Semiotica e filosofia del linguaggio*, Einaudi, Torino.
- (1993), *La ricerca della lingua perfetta nella cultura europea*, Laterza, Roma–Bari.
- FODOR, J. (1983), *The Modularity of Mind: An Essay on Faculty Psychology*, MIT Press, Boston.
- (1994), *The Elm and the Expert, Mentalese and its Semantics*, MIT Press, Boston.
- (2000), *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, MIT Press, Boston.
- HOFSTADTER, D. *I am a strange loop*, Basic Books, New York.
- SNOW C. (1959), *Le due culture*, Feltrinelli, Milano.
- SPENCER H. (1864), *Principles of Biology, Volume 1*, Williams and Norgate, London.
- VAPNIK V. (1998), *Statistical Learning Theory*, Wiley–Interscience, Hoboken N.J.
- VOLLI U. (1993), *I filosofi e il linguaggio*, Esculapio, Bologna.
- (2002), *Figure del desiderio*, Raffaello Cortina, Milano.
- VOLLI U., ROBIATI BENDAUD V. (2021) *Discutere in nome del cielo*, Guerini, Milano.
- YATES F. (1966). *The art of memory*, London: Routledge.

IA E SEMIOTICHE APPLICATE

VERSO UNA SEMANTICA QUANTISTICA? APPLICAZIONI AL DISCORSO RELIGIOSO⁽¹⁾

FRANCESCO GALOFARO, ZENO TOFFANO*

ENGLISH TITLE: *Toward a Quantum Semantics? Applications to Religious Discourse.*

ABSTRACT: since 2014 the authors have been developing an algorithm which exports to the field of Information Retrieval some concepts and models from quantum theory: entanglement and correlation due to quantum interference. In this article we explore the linguistic and semiotic implications of a quantum-like semantic theory, reconstructing the relationships between our work and a wide field of study borrowing statistical instruments from quantum theory and applying them to humanities and social sciences. Finally, as an example, we will show how our algorithm works on a case study: religious discourse in new media. In particular, the application of quantum instruments is helpful to understand how the quotation of a Twitter message in the reply modifies its original semantic profile.

KEYWORDS: quantum-like, structural semantics, von Neumann information, non-commutativity, non-compositionality

A partire dal 2014 abbiamo proposto e sviluppato un algoritmo che esporta al campo dell'*Information Retrieval* alcuni concetti e modelli provenienti dal mondo dei quanti: in particolare, *entanglement* e correlazione. In questo articolo vorremmo esplorare i fondamenti linguistici e semiotici del nostro lavoro, in relazione a un più ampio ambito di studi che, data la natura statistica della fisica dei quanti, ne prende a prestito alcuni strumenti

(1) Questo progetto ha ricevuto finanziamenti dall'European Research Council (ERC) nell'ambito del programma di ricerca e innovazione Horizon 2020 dell'Unione europea, in virtù della convenzione di finanziamento n. 757314.

* Università IULM di Milano; Centrale Supélec.

applicandoli alle scienze umane e sociali. Infine, a titolo di esempio, mostreremo un'applicazione del nostro algoritmo a un caso studio specifico: la circolazione del discorso religioso nei nuovi media. In particolare, l'algoritmo è utile a comprender come l'atto di citare un messaggio di Twitter in una risposta ne modifichi il profilo semantico originario.

1. *Quantum-like e Quantum Information Retrieval*

L'*Information Retrieval* è stato il primo campo in cui si è delineata la possibilità di impiegare strumenti quantistici per acquisire informazioni sul piano del contenuto della lingua. Il tentativo si inserisce in un più ampio filone di ricerca sorto negli ultimi venti anni e noto come *quantum-like*:

Gli anni recenti sono caratterizzati da una vera esplosione dell'interesse alle applicazioni della teoria della misurazione quantistica e altre parti del formalismo quantistico al di fuori della fisica, in particolare in psicologia, decision making, economia, finanza e scienze sociali (Khrennikov 2020).

Come si vede dall'elenco, sono soprattutto le scienze sociali ad essere interessate dai modelli quantistici. In questo ambito, Zeno Toffano ha proposto diversi contributi proponendo una logica basata sugli operatori (Toffano e Dubois 2020) e una sua applicazione ai robot di Braitenberg (Toffano e Dubois 2019) che presenta un interesse semiotico specie per quanto riguarda la formalizzazione delle passioni.

In via generale, l'*Information Retrieval* si dedica a estrarre informazione dai documenti. In questo ambito, quello di Keith van Rijsbergen (2004) è il volume pionieristico che ha aperto alle applicazioni della matematica dei quanti. Lo scopo per cui, in origine, Van Rijsbergen ha proposto l'adozione del formalismo dei quanti è molto pratico: unificare i tre approcci tradizionali, paralleli e non sempre comunicanti, al problema dell'*Information Retrieval*: logico, geometrico e probabilistico.

Questi tre e probabilmente anche altri possono essere descritti nello spazio di Hilbert. Il ragionamento che si trova in ciascuno di questi modelli

è formulato algebricamente e si può mostrare come esso dipenda essenzialmente dalla geometria dello spazio dell'informazione. La geometria può essere vista come un linguaggio per esprimere i differenti modelli di Information Retrieval (Rijsbergen 2004, p. ix, traduzione nostra).

Nella proposta di van Rijsbergen, ogni documento è rappresentato come in fig. 1: un *vettore unitario* nello spazio dell'informazione generato da una base (X, X_{\perp}) che rappresenta la presenza/assenza di una certa proprietà, ad esempio la pertinenza (*relevance*) a una domanda di ricerca (*query*). Le proiezioni del vettore-documento sugli assi rappresentano le *ampiezze di probabilità* che esso sia pertinente o meno alla domanda, mentre sono i rispettivi *quadrati* delle ampiezze a darci le probabilità che il documento risulti pertinente o non pertinente. Per il teorema di Pitagora, infatti, la somma delle probabilità è sempre 1. In questo modello, i valori massimali sono infatti 1 e 0 e rappresentano la certezza che un certo documento sia pertinente o meno al quesito di ricerca. I valori di probabilità così ottenuti permettono a un motore di ricerca di classificare i risultati in base alla pertinenza presunta. Nell'ottica presentata qui, il documento è rappresentato come un vettore che ruota nello spazio a seconda delle operazioni che compiamo su di esso: nella misura in cui muta l'angolo del vettore con le proprie basi, cambiano anche le ampiezze di probabilità relative.

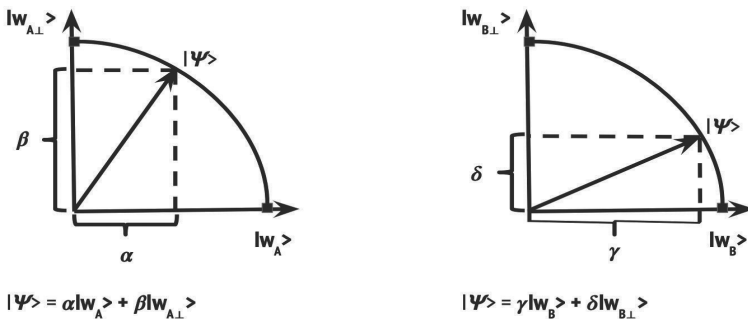


Figura 1. Rappresentazione geometrica del documento come un vettore unitario $|\Psi\rangle$ in uno spazio di Hilbert. Il medesimo documento può essere rappresentato in basi diverse, ad esempio $\{A, A_{\perp}\}$ e $\{B, B_{\perp}\}$; ciascuna base rappresenta la presenza o l'assenza di una certa proprietà semantica.

Da un punto di vista semiotico, questo modello sembra utile per definire operativamente ciò che si intende come *pertinenza* a una domanda di ricerca nei termini di possesso di una certa proprietà semantica da parte del documento; sappiamo che qualunque messaggio vive in uno spazio semantico che è sempre contraddittorio (Eco 1975) e la teoria quantistica fornisce un mezzo per rappresentarlo; una misura della probabilità della presenza di un certo livello di lettura coerente in un documento può rappresentare il *peso* di una data *interpretazione*.

Dagli studi di van Rijsbergen è nato un fertile campo di studi inteso a verificare l'esportabilità di frammenti di teoria quantistica dalla fisica all'*Information Retrieval*. Una rassegna molto completa dei primi dieci anni di vita della disciplina si deve a Massimo Melucci. Tra gli altri concetti che possono rivelarsi utili all'*Information Retrieval*, Melucci annovera la nozione di *entanglement* per rendere conto della *correlazione* e fornire un modello della "combinazione di concetti", richiama i lavori pionieristici in questo campo e indica nell'*entanglement* uno dei possibili campi di ricerca per il futuro (Melucci 2015, pp. 130, 145-146, 190-191). Su questo tema si veda anche Bruza, Kitto, Ramm, Sitbon, Song e Blomberg (2012). Non stupisce che vi sia una forte continuità di autori e metodi tra *Information Retrieval* quantistico e il più vasto campo della computazione quantistica (Melucci 2015, pp. 111, 141).

2. Geometria dello spazio semiotico

Veniamo ora ai motivi per cui la semiotica, tra le altre scienze sociali, dovrebbe interessarsi a propria volta ai modelli quantistici o *quantum-like*. L'approccio quantistico al problema del significato richiede una visione geometrica della semantica. Ciascun lessema o lessia è rappresentato come un vettore che vive nello spazio semantico del documento; la dimensionalità dello spazio coincide, almeno al principio, col numero dei lessemi in esso contenuti⁽²⁾; inoltre, come vedremo in 3.2, la natura

(2) La trasformazione del documento nello spazio dei vettori-parola avviene attraverso il metodo *Hyperspace Analog to Language* che trova largo impiego nel campo del quantum Information Retrieval (Melucci 2015, pp. 141-145). A differenza di altri metodi largo impiego, come *bag-of-words*, esso permette infatti di ottenere facilmente il genere di matrici cui si applica la teoria dei quanti: simmetriche, ortonormali, hermitiane (Susskind e Friedman 2014, pp. 48-51).

non compositazionale del significato spinge a dire che lo spazio da considerare è isomorfo a quello dei numeri complessi.

A questo proposito, la scelta di considerare strumenti quantistici di analisi non solo non si discosta dalla ricerca semiotica, ma è in linea con una tradizione di ricerca ben consolidata (René Thom, Jean Petitot, Wolfgang Wildgen, Alessandro Sarti, per citare solo alcuni autori). Nel corso della storia ultrasecolare della nostra disciplina, infatti, diversi studiosi hanno discusso e proposto modelli per spiegare il funzionamento del sistema presupposto dalle strutture sintagmatiche che essi intendevano spiegare. Se in Saussure il sistema prendeva la forma di catene associative (*grandemente*, *intensamente*, *indelebilmente* ...), esso appariva a Hjelmslev e Jakobson come un insieme di opposizioni logiche caratterizzate dall'operatore aut ("il (/gatto/ aut /cane/ aut /topo/) mangia il formaggio"). Con Chomsky e Greimas l'algebra del sistema acquista la complessità dell'algoritmo⁽³⁾. Jean Petitot (1979, p. 103) ha mosso alla concezione greimasiana alcune critiche pertinenti al nostro problema:

- a) Come tutte le *descrizioni* logico-combinatorie, [scil. il modello di Greimas] non può rendere conto della *produzione* delle strutture, né della loro *dinamica*, né della loro *stabilità* (pregnanza). Riconduce in ambito narrativo l'*ideologia* di una grammatica categoriale.
- b) Forclude a priori la questione del soggetto, cioè il rapporto tra livello simbolico e semiotico nel senso di J. Kristeva.
- c) Nella sua regressione dalla struttura attanziale di superficie alla struttura semica profonda ricodifica sotto forma di *relazione* (di opposizione tra semi) quel che nei fatti è un *evento* conflittuale codificato da un *verbo*. Da una parte, riproduce dunque il pregiudizio logicista fondamentale dell'esclusione *a priori* del verbo e dell'evento che, pertanto, *organizza* la struttura e, d'altra parte, non permette di rendere conto delle diverse rappresentazioni *geometriche* del conflitto.
- d) Infine, non riconosce che certe strutture narrative mettono sempre in gioco degli eventi *d'altro livello*, nello specifico *eventi somatici*, (evidentemente ineliminabili nelle rappresentazioni pittoriche). Si situa a un solo livello (sintattico) dimenticando che una struttura, quale che sia, si mantiene attraverso un gioco dialettico di dispiegamento-ridispiegamento tra *due* livelli (nel nostro caso, sintattico/somatico) (1979, p. 103, traduzione nostra)

(3) Cfr. "Algoritmo" in Greimas e Courtés (1979).

Alcune critiche di Petitot risentono evidentemente del dibattito semiotico dell'epoca; altre sono motivate dal fatto che il suo saggio è dedicato al visivo; altre ancora si direbbero più generali e attuali: il modello greimasiano non rende conto della produzione delle strutture che descrive, della loro dinamica o della loro stabilità; inoltre, ogni struttura si mantiene attraverso un gioco dialettico di dispiegamento–ridispiegamento tra *due* livelli. Per queste ragioni, Petitot adotta la teoria delle catastrofi allo scopo di proporre modelli i quali “determinano le soluzioni elementari stabili al problema della competizione spaziale retta da un sistema di soglie e sviluppano un evento organizzatore, produttore e strutturante”. In particolare, egli è alla ricerca di un modello che generi quel che per Greimas è la *sintassi del contenuto*. Rinvio a René Thom (1985) per una panoramica sui tipi di interazione che è possibile generare attraverso la teoria delle catastrofi. Thom (2010) mostra come si possano generare pregnanze singolari a partire da un'unica gravidanza generale, ottenendo il noto quadrato semiotico di Greimas come risultato. Il modello di Thom è coerente e semplice, presentando il vantaggio di spiegare e dispiegare la distinzione greimasiana tra livello fondamentale e semionarrativo generando entrambi con un solo gesto elegante. Il legame tra lo spazio di partenza della gravidanza generale e quello di arrivo dove si collocano le pregnanze individuali è visto come un processo di *individuazione*.

Thom riprende le nozioni di individuazione e di *metastabilità* dall'epistemologia di Gilbert Simondon (2005). Egli definisce l'informazione come un processo che porta all'individuazione di una forma a partire da uno stato metastabile del sistema. A propria volta, l'individuazione è un fenomeno di risonanza e riconfigurazione interna del sistema a partire dalle differenze eterogenee pre-individuali che ne fanno parte. L'integrazione delle risonanze genera differenze globali. Secondo Simondon, tale processo è sempre incompleto. La forma è un network di differenze pre-individuali e di caratteristiche individuali. L'individuo non è mai una monade isolata, ma porta inscritta in sé una dimensione collettiva, trans-individuale. In termini matematici, l'emergere di una forma da uno spazio si ottiene attraverso il calcolo delle autofunzioni di un grafo immanente (Sarti e Piotrowski 2015, p. 57). Per comprendere il passaggio alla matematica dei quanti, è importante evidenziare che,

in un modello stocastico, si può descrivere tale grafo in termini di densità di probabilità (Sarti e Piotrowski 2015, p. 56).

Una differenza importante tra Thom e Simondon riguarda la rispettiva posizione sul determinismo. La matematica di Thom è sempre, per scelta, deterministica. Al contrario, secondo Simondon, è proprio l'incompiutezza del processo di individuazione a impedire di renderne conto nei termini della logica classica, ovvero del principio di identità o del terzo escluso. Per questo, Simondon cerca un'alternativa nella teoria dei quanti, la quale, a partire dal concetto di *sovrapposizione coerente di stati contraddittori*, sembra possedere i mezzi per far presa su questa materia sdruciolevole. Su questa linea di ricerca, Francesco Galofaro (2015) ha mostrato come sia possibile rappresentare in termini di circuiti quantistici quelle interazioni per generare le quali Thom aveva utilizzato la teoria delle catastrofi. Inoltre, la computazione quantistica permette di formalizzare la nozione greimasiana di comunicazione partecipativa (Greimas 1983, pp. 40–42), in cui un oggetto di valore è trasferito tra due attanti senza che l'attante-fonte se ne privi, come avviene nel processo di apprendimento o nella delega dei poteri in politica.

3. Adeguatezza dei modelli quantistici

La preferenza verso i modelli quantistici non è ovviamente fine a se stessa. Questa particolare matematica pare adeguata ad alcune caratteristiche presentate dal nostro oggetto di studio. Abbiamo sviluppato le implicazioni epistemologiche dell'adozione di strumenti quantistici per quanto riguarda la linguistica e la semiotica in Galofaro, Doan e Toffano (2016). Qui ci limitiamo a riassumere le questioni principali.

3.1. Non commutatività del significato

Nell'*Information Retrieval* la non commutatività è una proprietà molto importante. Se si effettuano due ricerche entro un corpus, l'ordine in cui le si esegue è rilevante. Se si selezionano *prima* tutti i documenti che contengono la parola "rock" (*aboutness*) e *poi*, tra questi, quelli

pertinenti alla musica (*relevance*), si ottiene un insieme di risultati; viceversa, anteponendo la *relevance all'aboutness*, se ne otterranno altri. Così è anche in semantica, dove l'ordine delle parole conta: le frasi "I marxisti criticano gli economisti liberali" e "Gli economisti liberali criticano i marxisti" esprimono significati opposti; tuttavia, una ricerca come /marxismo/ + /economia/ restituirà indifferentemente risultati pertinenti a entrambi i significati. Al contrario, a differenza degli operatori della logica classica, l'ordine in cui si presentano gli operatori nella computazione quantistica cambia il risultato, riflettendo in questo modo l'ordine del linguaggio.

3.2. *Non composizionalità del significato*

Ci si può chiedere perché lo spazio del significato debba essere uno spazio di Hilbert⁽⁴⁾. La risposta è che, in uno spazio di Hilbert complesso, sono possibili *ampiezze di probabilità negative*. La loro funzione è quella di rendere possibile una "interferenza" tra le probabilità, cosa che la nozione di probabilità in sé non permetterebbe (non riusciamo ad associare nulla di preciso a una "probabilità negativa"). In termini semantici, dunque, i significati dei lessemi possono *interferire tra loro*, e il senso complessivo di un messaggio non è una mera somma dei significati degli elementi semplici (non composizionalità del significato).

Per esempio, la lingua contiene molte espressioni non composizionali, chiamate *lessie* nella tradizione della linguistica strutturale. Pensiamo alle espressioni idiomatiche, il cui significato non può essere considerato come una sommatoria di elementi. Ad esempio, l'espressione polacca *musztarda po obiedzie* (senape dopo pranzo) risulta piuttosto oscura a chi è esterno alla lingua e alla cultura di appartenenza: si impiega per qualcosa che sarebbe stato utile ma è giunto troppo tardi. La non composizionalità del linguaggio diviene un problema importante per la *sentiment analysis* e, più in genere, per il vasto campo dei *Big Data*, come conferma un recente studio:

(4) La domanda è stata posta a Francesco Galofaro da Jean Petitot, cui vanno i ringraziamenti degli autori.

Examples, like “break a leg,” “speak of the devil,” and “let the cat out of the bag,” would imply the wrong sentiment based on the negative sentiment in “break” and “devil” and lack of strong polar sentiment in any of the words “let,” “the,” “cat,” “out,” “of,” and “bag.” (Hwang e Hidely 2019).

In termini semiotici, il “sentiment” sembra essere parte del piano del contenuto del messaggio: quel che accade è che un valore forico ipercodificato nella cultura in associazione a un lessema come /devil/ non corrisponde a quello associato alla lessia in cui la parola “devil” incorre.

In realtà, i casi di non composizionalità vanno molto oltre il campo delle espressioni idiomatiche: in una certa misura, per piccola che sia, quanto ogni messaggio rinegozia, innova, talvolta sovverte il significato che la cultura associa ai propri termini. Ad esempio, per il dizionario /sabbia/ è un iponimo di /roccia/, ma in Luca 6, 47–49 i due termini divengono antonimi, per omologazione, a /saggio/ e /stolto/, perché il testo seleziona un’opposizione privativa che li distingue (potere/non poter fondare una casa).

3.3. *Non distributività delle relazioni di iponimia*

Nel formato enciclopedico della biologia, l’intersezione tra /umani/ e /uccelli/ è vuota: nella nostra enciclopedia, non vi sono uomini–uccelli se non in testi a carattere finzionale quali i fumetti dei supereroi, che manifestano una visione totemica del mondo. Lo stesso si può dire dell’intersezione tra /uccelli/ e /lucertole/. Ma se consideriamo l’unione tra /umani/ e /lucertole/, otteniamo la classe dei /vertebrati/. E l’intersezione tra /vertebrati/ e /uccelli/ non è vuota, ma è costituita, di nuovo, dagli /uccelli/. Segue che la relazione di iponimia — se /uccelli/ allora /vertebrati/ — non è distributiva:

$$x \cap (y \cup z) \neq (x \cap y) \cup (x \cap z)$$

Questo assioma non è valido nella logica classica ma lo è, ancora una volta, in quella quantistica (Dalla Chiara e Giuntini 2002, cap. 1)

3.4. *Fallimento del principio di sostitutività*

Un vecchio principio che si deve a Frege vuole che, se due parole possono essere sostituite in tutti i contesti, esse hanno lo stesso significato: è un adattamento alla lingua della legge di Leibniz dell'*indiscernibilità degli identici* per cui due enti sono uguali se ogni predicato può essere detto di entrambi. Tale assioma non sembra valido in ogni contesto linguistico. Ad esempio, l'inglese ha due parole sostituibili in ogni contesto: *isle* e *island*. Esse, tuttavia, non hanno il medesimo senso entro il discorso etimologico. Infatti, il primo deriva dal latino *insula*, riconducibile a *in salo*, ovvero “nel mare”, il secondo dall'antico inglese **ieg-land*, che potremmo rendere con “acqua-terra”. Dunque, il principio funziona in contesti estensionali — quelli in cui la funzione prevalente è quella referenziale — ma non in contesti intensionali. Come accade per il principio di sostitutività, anche l'abolizione della legge di Leibniz ci fa uscire dall'ambito della logica classica per entrare in quella quantistica (Takeuti 1981).

3.5. *Correlazione ed entanglement*

La misura del grado di correlazione e di *entanglement* tra i vettori di base che compongono lo spazio semantico del documento è il focus principale delle nostre ricerche. Originariamente, la correlazione è legata alla misura dei fenomeni di *interferenza* tra particelle. Esse hanno trovato impiego nell'*Information Retrieval* (Galofaro, Toffano e Doan 2017) e nell'analisi automatica del contenuto dei messaggi pubblicati su social network (Galofaro e Toffano 2020). In questa occasione vorremmo riflettere meglio sul significato dell'*entanglement* nei termini della semantica strutturale. In effetti, la nozione di *entanglement* è riferita alla non-separabilità di un sistema in sottosistemi (Susskind e Friedman 2014, p. 117; Melucci 2015, pp. 111–113). In riferimento alla fig. 1, uno spazio semantico S_{AB} è fattorizzabile in due sottosistemi S_A e S_B se è uno *stato prodotto*:

$$S_{AB} = S_A \otimes S_B$$

In caso contrario, lo spazio semantico non è ulteriormente analizzabile in sottospazi indipendenti e gli elementi dello spazio semantico si dicono *entangled* (“intricati”). La presenza di questo legame tra due lessemi ci dice, insomma, che essi fanno parte della medesima *isotopia*, della stessa distribuzione coerente di valori lungo la dimensione sintagmatica del testo; oppure, nel caso che tra i due lessemi si dia una relazione di anticorrelazione, essi fanno parte di un’*allotopia*; in entrambi i casi non è possibile considerare indipendenti i due spazi semantici che i due lessemi generano in quanto basi.

3.6. La nozione-ponte: informazione quantistica

L’adeguatezza non è l’unico criterio che ci permette di applicare la matematica dei quanti alla semantica strutturale. Non si tratta di dichiarare “basta che funzioni”. Tra matematica e semantica vi è una nozione-ponte che permette di comprendere meglio il risultato delle operazioni dei ricercatori nell’uno e nell’altro campo: la nozione di *informazione*. In ambito quantistico, tutte le informazioni che un osservatore ha sul sistema possono essere codificate in un operatore ρ , il quale, scelta una base (X, X_{\perp}) ci restituisce una matrice densità:

$$\rho_{XX_{\perp}} = \langle X | \rho | X_{\perp} \rangle$$

che comprende un valore di probabilità per ciascuno degli stati che il sistema può assumere (Susskind e Friedman 2014, pp. 152–153). Occupiamoci ora della differenza tra il caso di un sistema classico e un sistema *entangled*:

Uno stato puro classico è un caso particolare di una densità di probabilità in cui ρ è diversa da 0 solo in un punto (...). Se ρ è distribuita, vuol dire che la nostra conoscenza dello stato del sistema è incompleta: più è distribuita, maggiore è la nostra ignoranza (Susskind e Friedman 2014, p. 154)

In un sistema classico, se conosciamo lo stato puro per il sistema combinato di due elementi, allora conosciamo tutto di ciascun elemento,

che si tratti di particelle oppure — nel nostro caso — di lessemi. “Ma questo è esattamente ciò che non è vero (...) nel caso di un sistema *entangled*. Lo stato di un sistema composto può essere puro, ma i suoi costituenti devono essere descritti da una miscela (Ivi)” ovvero da una matrice ρ distribuita.

Indipendentemente dal fatto che il sistema comprendente i lessemi delle basi che ci interessano sia uno stato prodotto oppure *entangled*, quel che otteniamo applicando il nostro algoritmo è ottenere *informazione* sulla semantica di un messaggio, ovvero sulla *forma del suo contenuto*. Occorre prestare attenzione a questo punto, perché non otteniamo il *senso* del documento: il nostro algoritmo non fa l’*esegesi* del messaggio, non ne propone un’interpretazione, ma restituisce *informazione sul senso*: ad esempio, “nel documento y i lessemi l_1 e l_2 sono utilizzati come antonimi”.

4. Un esempio di applicazione al discorso cattolico ecologista

La pubblicazione della lettera enciclica di papa Francesco *Laudato si* è stata percepita come una svolta nel discorso cattolico sull’ambiente. In una nota Bruno Latour l’ha definita “sorprendente”. Secondo Latour,

the Christians, having lost the race for the most indisputable type of certitude, have gradually abandoned all concern with the cosmos in order to devote themselves to the salvation of humans alone, and then among the humans to the salvation of the soul alone, before abandoning the soul itself to the exclusive benefit of morality (Latour 2015, p. 210).

Dunque, perso interesse per il destino del Cosmo, i cristiani hanno cominciato a considerare l’ecologia come una forma di paganesimo. La crisi ecologica avrebbe pertanto radici religiose. A maggior ragione, dunque, *Laudato si* è un documento interessante per comprendere come cambia nel discorso sociale il rapporto tra religione e ambiente e il rapporto tra cosmologia e senso del sacro. In quel che segue, vorremmo mostrare come le tecniche di semantica quantistica si prestino a indagarne la circolazione e l’ibridazione col concetto di antropocene. Nozione secolare, ma — come vedremo — solo in superficie,

quest'ultima non compare nel testo di papa Francesco, mentre nei tweet finisce per fondersi con l'assiologia cattolica tramite la giustapposizione di due hashtag (#anthropocene e #LaudatoSi) e la sovrapposizione tra discorso politico e religioso.

4.1. *Corpus*

Per rispondere alla domanda, abbiamo raccolto un corpus di tweet contenente l'hashtag #LaudatoSi, dall'anno 2015 ad oggi. L'insieme comprende 2164 commenti in inglese. Per studiare le caratteristiche semantiche dei tweet che associano gli hashtag #laudatosi, #sustainable e #anthropocene abbiamo selezionato un sottoinsieme di commenti, che ritwittavano, citandolo, un tweet originale riferito all'enciclica, aggiungendo un secondo hashtag.

È in primo luogo interessante notare che nel nostro corpus non vi siano tweet che associno all'hashtag #LaudatoSi tanto l'hashtag #anthropocene quanto gli hashtag #sustainable, #sustainability o #sustainabledevelopment. Ciò è coerente col fatto che, pur non utilizzando mai il termine “antropocene”, l'account Twitter ufficiale di Papa Francesco @Pontifex utilizza spesso l'aggettivo “sostenibile”. Questa interessante disgiunzione riflette anche il fatto che la parola “antropocene” non ricorre mai nell'enciclica, mentre l'aggettivo “sostenibile” ricorre nove volte in associazione con “sviluppo”, “uso”, “agricoltura”, “progresso”, ed è presente una sola volta nella forma negativa “insostenibile”.

Dunque, empiricamente, i due temi (sostenibilità e antropocene) rappresentano un'alternativa paradigmatica: #LaudatoSi è una *costante* che può essere selezionata sia dalla *variabile* #anthropocene sia dalla *variabile* #sustainable. Ribadiamo che questa disgiunzione paradigmatica non è pertinente all'universo dei tweet a tema ecologico; essa caratterizza soltanto quei tweet ecologici che includono l'hashtag #LaudatoSi, contribuendo in qualche modo a definirli.

4.2. *Metodologia*

L'associazione tra la coppia di hashtag #LaudatoSi e #anthropocene è stata analizzata applicando l'algoritmo semantico quantistico

sviluppato da Galofaro, Toffano e Doan (2020). Tale algoritmo permette di visualizzare lo spazio semantico del messaggio originale e le trasformazioni geometriche che in esso si generano quando il messaggio viene citato e ritwittato insieme a una risposta. L'algoritmo restituisce una rappresentazione dello spazio semantico del documento nei termini dei due valori di correlazione e di *entanglement* che ciascun lessema totalizza nei confronti di tutti gli altri. Per far questo, il documento è preliminarmente trasformato in un vettore ed espresso nei termini di ciascun vettore-lessema di base X e del suo vettore ortogonale X_{\perp} , come nella fig. 1 che abbiamo già commentato al par. 1. Considerando la loro co-occorrenza nei diversi contesti, i lessemi del documento possono essere:

- correlati, quando tendono a co-occorrere negli stessi contesti: il punteggio di correlazione tende al valore $+1$;
- anticorrelati, quando tendono a non co-occorrere negli stessi contesti: il punteggio di correlazione tende a -1 ;
- non correlati, quando non co-occorrono con una precisa tendenza: in tal caso, il punteggio tende a 0 ;

Correlazione e anti-correlazione possono aiutare l'analista a identificare lessemi isotopici (appartenenti allo stesso livello di coerenza semantica) e lessemi allotopici (appartenenti a opposti livelli semantici).

Venendo all'*entanglement*, come si è detto, il suo valore misura il grado di inseparabilità di un dato sistema in sottosistemi semantici indipendenti. Esso può assumere un punteggio tra 0 e $2\sqrt{2}$ (*Tsirelson bound*). È considerato non-classico o *entangled* un valore superiore a 2 (disuguaglianza CSHS — cfr. Clauser, Horne, Shimony e Holt 1969).

La differenza principale tra *entanglement* e correlazione consiste nel fatto che tanto i lessemi allotopici quanto quelli isotopici possono appartenere allo stesso sottospazio semantico inseparabile: ciò pare coerente con il punto di vista della semantica strutturale, secondo la quale lo spazio semantico è formato da categorie (bene/male, maschio/femmina, vita /morte ...).

4.3. L'hashtag #anthropocene

Consideriamo il tweet (T):

T) My favorite part of Pope Francis' encyclical: A new prayer to help us commune with Nature (pubblicato da @EricHolthaus nel 2015)

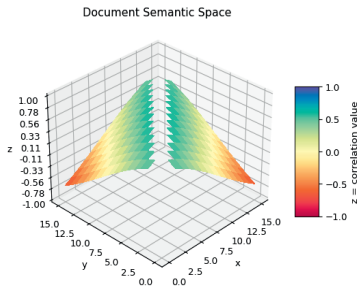
Il tweet accompagna uno screenshot di una delle due preghiere che concludono l'enciclica del papa. Il messaggio originale è citato nel seguente retweet (R):

R) "Prayer for California #wildfires, #PuertoRico, and the era we live in. From @Pontifex encyclical #ClimateAction #anthropocene" (pubblicato da @DianeDulken nel 2017)

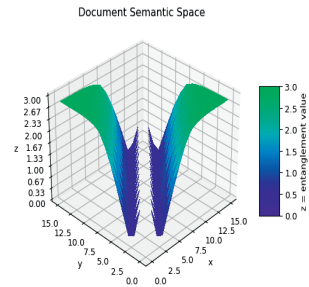
Il modo in cui R modifica lo spazio semantico di T è rappresentato in figura 2. Gli assi x e y rappresentano gli n lessemi appartenenti al messaggio. Per ogni coppia di lessemi, l'asse z rappresenta il loro valore di (anti-) correlazione (figg. 2a e 2c) o *entanglement* (figg. 2b e 2d). È facile apprezzare, da un punto di vista qualitativo, la superficie "liscia" dello spazio semantico del tweet originale: essa è dovuta al fatto che ogni lessema ricorre una sola volta nel testo. In questa situazione, il grado di correlazione dipende banalmente dalla distanza tra i lessemi nella catena sintagmatica. La superficie si fa più increspata nel retweet, in cui le relazioni semantiche sono più complesse per via del ricorrere di alcuni lessemi in diversi contesti del documento. Naturalmente, più ampio è il testo e più questo genere di rappresentazioni spaziali si fa complesso e interessante.

Nel caso in esame è utile studiare una specifica sezione dello spazio corrispondente a tutti i valori di correlazione ed *entanglement* di un certo lessema di nostro interesse. Ad esempio, partendo dagli spazi semantici in figg. 2a e 2c, rappresentiamo in fig. 3 la sezione dello spazio corrispondente al lessema /nature/. Il retweet della citazione contenente l'hashtag #anthropocene rimodella i valori originari, rafforzandoli: si veda ad esempio il nesso semantico che viene a crearsi tra /Nature/ ed /encyclical/. Si formano nuove interessanti correlazioni, ad esempio tra /Nature/ e /life/.

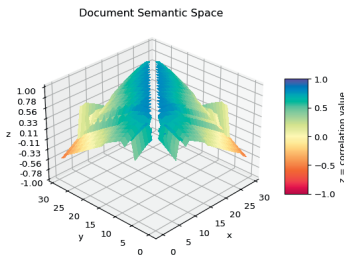
a) Correlazione nel tweet originale



b) entanglement nel tweet originale



c) Correlazione nel retweet



d) entanglement nel retweet

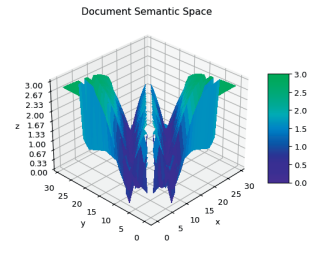
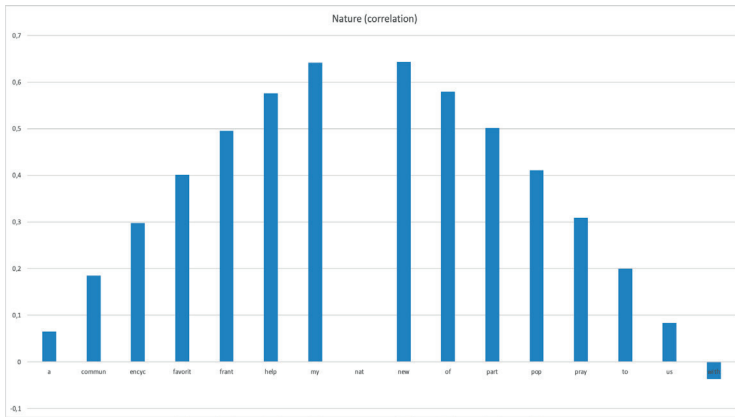


Figura 2. Come il retweet R modifica lo spazio semantico del tweet C.

I valori di entanglement in fig. 4 sono più pertinenti al quesito della nostra ricerca. I valori superiori a 2 indicano la presenza di *entanglement*. Nel tweet originale il lessema /nature/ è intrecciato con il lessema /us/ (fig. 4a). Questo è coerente con il significato del primo tweet, stabilendo una comunione tra noi e la natura. Nel retweet (fig. 4b) /nature/ si intreccia solo con /anthropocene/. Quindi, nel retweet, /anthropocene/ e /nature/ non possono essere considerati parte di diversi sottosistemi isotopici: i loro significati sono assimilati.

Veniamo ora alla sezione dello spazio semantico corrispondente al lessema /anthropocene/: i valori di correlazione ed entanglement sono riportati in fig. 5. Considerando i lessemi del tweet originale, il lessema antropocene stabilisce la correlazione più forte con /commune/, /Francis/, /Nature/ (fig. 5a) e le più forti relazioni di entanglement con /Nature/, /Pope/ e /prayer/ (fig. 5b), formando con essi un sistema inseparabile. In questo modo, valori sacri, in qualche modo mistici,

a) tweet originale



b) retweet

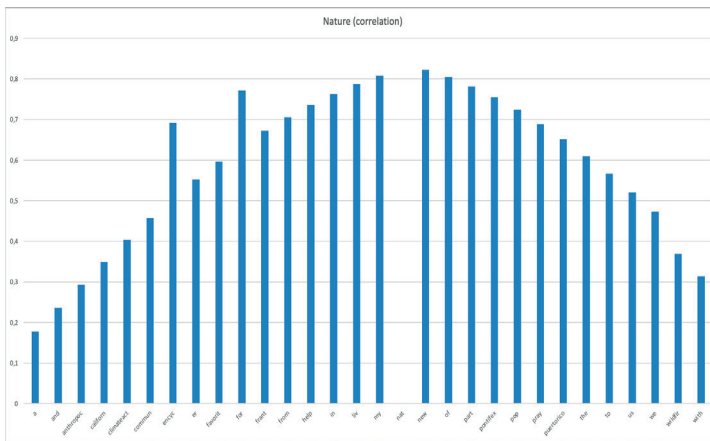


Figura 3. Il retweet contenente l’hashtag #anthropocene modifica i valori di correlazione del lessema “nature”: (a) il tweet citato; (b) il retweet che lo cita.

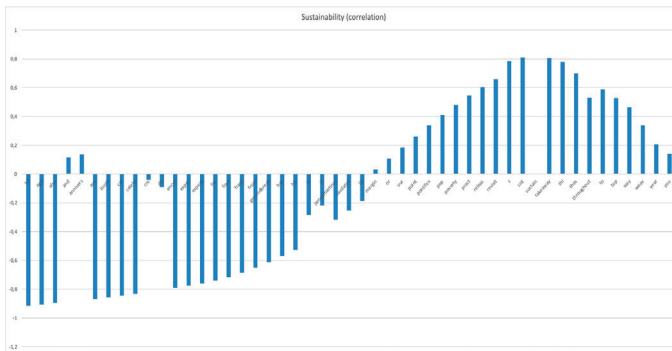
vengono trasferiti al lessema /anthropocene/ a partire da questi tre termini, che se ne fanno vascello. Allo stesso tempo, questa sfumatura “mistica” fornita dal rapporto tra natura e preghiera, con l’intervento di papa Francesco, è creata proprio dall’apparizione dell’antropocene nel sistema semantico, poiché questi lessemi non erano *entangled* nel tweet originale (fig. 4a).

10 takeaways from #LaudatoSi (pubblicato da @Americamag in 2019)
 R) Care for Creation is also care for our brothers and sisters, especially those experiencing poverty or marginalization. Pope Francis @Pontifex weaves this point throughout #LaudatoSi. Do you have any ways you like to practice #sustainability? (pubblicato da @sColumbkille nel 2019)

In questo caso, il lessema /sustainability/ è *entangled* con /takeaways/ e il lessema /celebrate/, appartenenti al messaggio originale (fig. 6b).

/Takeaways/ e /celebrate/ erano già *entangled* nel messaggio originale con un punteggio molto alto (2,77). Tuttavia, mentre /sustainability/ è positivamente correlata con /takeaways/, il legame con il /celebrate/ dà

a) valori di correlazione



b) valori di entanglement

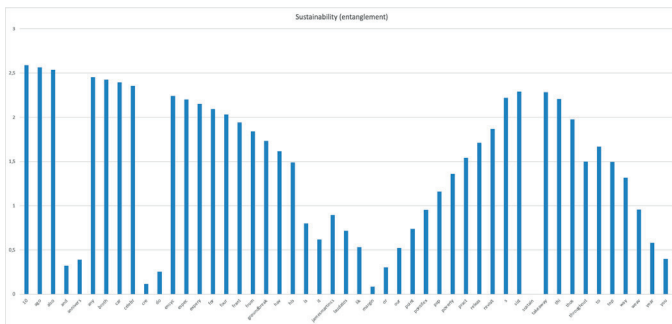


Figura 6. Colpo d'occhio sui valori di correlazione ed entanglement del lessema /sustainability/ con tutti gli altri lessemi.

luogo qui a una forte anticorrelazione ($-0,83$). I lessemi /Sustainability/ e /celebrate/ appartengono sì allo stesso sistema, manifestano una stessa categoria, ma possono essere considerati allotopici, ossia resi antonimici dal retweet. Come nel caso precedente, abbiamo a che fare con un tweet euforico citato in un retweet disforico, il cui focus principale è su povertà ed emarginazione. Confrontando questo retweet con quello che include l'hashtag #anthropocene, la differenza principale sembra consistere nel fatto che #anthropocene era associato a /prayer/ e a /comune/, mentre la sostenibilità non è correlata a nessun valore che manifesti il sacro e non è dunque ascrivibile al discorso religioso. Questa assenza coinvolge anche il tweet originale di American Magazine, che diffonde importanti messaggi ecologici e sociali nel format dei suggerimenti di moda.

5. Conclusione

L'enciclica *Laudato si* aggiorna il linguaggio poetico di san Francesco per proporre una nuova prospettiva francescana sull'ambiente. Trasferisce valori sacri al discorso ecologico e fornisce un modello, sotto forma di teologia politica, per cambiare la società e il potere. In questo quadro, i nostri strumenti hanno mostrato come il nuovo discorso cattolico circoli e si fonda con altri temi e figure del discorso ecologico, come la sostenibilità e l'antropocene. Abbiamo notato che, nel sottoinsieme dei discorsi ecologici che si riferiscono all'enciclica *Laudato si*, i temi "antropocene" e "sviluppo sostenibile" sono alternativi (cioè paradigmaticamente disgiunti). Inoltre, i sostenitori dell'antropocene hanno sollevato dubbi sul fatto che lo sviluppo sostenibile sia solo un mito "come le ideologie, che portano solo a posporre il problema" (Zinna 2015, p. VII). Da un punto di vista semiotico, tuttavia, sia l'antropocene sia la sostenibilità sono miti in senso tecnico, senza connotazioni negative: la questione d'attualità politica è se essi siano sufficienti per innescare i cambiamenti ecologici e sociali auspicati da coloro che li impiegano.

La metodologia che abbiamo utilizzato permette di mostrare come l'associazione con l'uno o l'altro hashtag cambi profondamente i valori in gioco. Coloro che impiegano il termine antropocene lo associano

intenzionalmente a una coordinata temporale valorizzata disforicamente (solitamente, il presente) e ad alcuni valori sacri, lasciando intravedere una teologia politica apocalittica (Latour 2005). L'apocalisse è il tratto distintivo di un sottogenere del discorso ecologico radicale. Tale associazione tra discorso ecologico e apocalittico non implica necessariamente la presenza del tema della sostenibilità.

Dal canto suo, l'aggettivo "sostenibile" può essere associato ad ogni tipo di argomento particolare, aggiungendogli un valore "ecologico". Ad esempio, si veda il seguente tweet del nostro corpus:

Make this Christmas a sustainable one. #LaudatoSi #COP26
#ClimateAction #ClimateJustice #edmundspeople.

Anche nella lettera enciclica, così come nel *Cantico delle creature* di San Francesco, è assente ogni forma di millenarismo e di escatologia. Piuttosto, la sostenibilità consente all'enunciatore di ricorrere al bricolage come pratica di creazione di significato (Floch 1995). In altre parole, la nozione di sostenibilità non è apocalittica quanto quella di antropocene; essa si presenta piuttosto come l'etichetta di un insieme di tattiche ecologiche locali che sfidano la strategia complessiva del potere (De Certeau 1980), una collezione di zeppe per tenere aperta una porta verso il futuro.

Riferimenti bibliografici

- BRUZA P.D., KITTO K., RAMM B., SITBON L., SONG D., BLOMBERG S. (2012), "Quantum-like non-separability of concept combinations, emergent associates and abduction" in *Logic Journal of the IGPL*, vol. 20, no. 2, pp. 445-457, doi: 10.1093/jigpal/jzq049 (consultato l'8 maggio 2022).
- CLAUSER F., HORNE M.A., SHIMONY A., HOLT R.A. (1969), "Proposed experiment to test local hidden-variable theories", *Phys. Rev. Lett.*, 23 (15), pp. 880-4.
- DALLA CHIARA M.L., GIUNTINI R. (2002), "Quantum Logics", in *Handbook of Philosophical Logic*, vol. 6, Dordrecht, Kluwer Academic Publishers, pp. 129-228.

- DE CERTEAU M. (1980), *L'Invention du Quotidien. Vol. 1, Arts de Faire*, Union générale d'éditions, Paris, tr. it. *L'invenzione del quotidiano*, Edizioni Lavoro, Roma, 2001.
- ECO, U. (1975), *Trattato di semiotica generale*, Bompiani, Milano.
- FLOCH J.-M. (1995), *Identités visuelles*, Puf, Paris, tr. it. *Identità visive*, Franco Angeli, Milano, 1997.
- GALOFARO F. (2015), "Structural Syntax and Quantum computation: a Simondonian Approach", in SARTI, MONTANARI, GALOFARO (a cura di) *Morphogenesis and Individuation*, Springer, Berlin, 2015, pp. 173–201.
- GALOFARO F., DOAN B.-L., TOFFANO Z. (2016), "Linguistics and Quantum Theory: Epistemological Perspectives" in *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*, pp. 660–668, 10.1109/CSE-EUC-DCABES.2016.257.
- GALOFARO F., TOFFANO Z. (2020), "Semantic correlation in hate speeches", RIFL, doi: 10.4396/SFL2019ES08.
- GALOFARO F., TOFFANO Z. DOAN B.-L. (2017), "A quantum-based semiotic model for textual semantics" in *Kybernetes*, doi: 10.1108/K-05-2017-0187.
- GREIMAS A.J. (1983), *Du sens II*, Seuil, Paris, tr. it. *Del Senso 2*, Bompiani, Milano, 1984.
- GREIMAS A.J., COUIRTÉS J. (1979), *Sémiotique: dictionnaire raisonné de la théorie du langage*, Hachette, Paris (trad. it. P. FABBRI (a cura di) *Semiotica: dizionario ragionato della teoria del linguaggio*, Bruno Mondadori, Milano, 2007).
- HWANG A., HIDEY CH. (2019), "Confirming the Non-compositionality of Idioms for Sentiment Analysis", in *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, Association for Computational Linguistics, Firenze, pp. 125–129, <https://aclanthology.org/W19-5114/> (consultato il 6 maggio 2022).
- LATOUR B. (1998), "Petite philosophie de l'énonciation", in P.L. BASSO & L. CORRAIN (ed.) *Eloqui de senso. Dialoghi semiotici per Paolo Fabbri: Orizzonti, compiti e dialoghi della semiotica*, Milano, Costa & Nolan, pp.71–94.
- (2015), *Facing Gaia: Eight Lectures on the New Climatic Regime*, Cambridge, Polity Press (2017).

- MELUCCI M. (2015), *Introduction to Information Retrieval and Quantum Mechanics*, Berlin, Springer.
- PETITOT J. (1979), “Saint Georges: remarques sur l’espace pictural” Zeitoun, J. (a cura di), in *Sémiotique de l’espace*, Paris, Denoël Gonthier.
- SARTI A., PIOTROWSKI D. (2015), “Individuation and Semiogenesis: An Interplay Between Geometric Harmonics and Structural Morphodynamics”, in SARTI, MONTANARI, GALOFARO (a cura di) *Morphogenesis and Individuation*, Springer, Berlin, 2015, pp. 49–73.
- SIMONDON G. (2005), *L’individuation à la lumière des notions de formes et d’information*, Jérôme Millon, Grenoble.
- SUSSKIND L., FRIEDMAN A. (2014), *Quantum Mechanics: the theoretical minimum*, Penguin, New York, tr. it. *Meccanica quantistica: il minimo indispensabile per fare della (buona) fisica*, Raffaello Cortina Editore, Milano, 2015.
- TAKEUTI G. (1981), “Quantum Set Theory”, in *Current issues in quantum logic*, vol. 1, Plenum, New York, pp. 303–322.
- TOFFANO Z., DUBOIS F. (2019), “Quantum eigenlogic observables applied to the study of fuzzy behaviour of Braitenberg vehicle quantum robots”, *Kybernetes*, Vol. 48 No. 10, pp. 2307–2324, doi: 10.1108/K-11-2018-0603.
- (2020), “Adapting Logic to Physics: The Quantum–Like Eigenlogic Program.”, *Entropy* 22, doi: 10.3390/e22020139
- THOM R. (1985), “Topologia e significazione”, in *Modelli matematici della morfogenesi*, Einaudi, Torino, pp. 192–212, poi in P. FABBRI (a cura), *Morfologia del semiotico*, Meltemi, Roma, pp. 23–53, 2006.
- (2010), *Strutture cicliche in semiotica*, Documenti di lavoro e prepubblicazioni del CISSM, 393–394–395, Urbino.
- VAN RIJSBERGEN C.J. (2004), *The Geometry of Information Retrieval*, Cambridge University Press, Cambridge.
- ZINNA A. (2015), “Les visages de Gaïa”, in A. ZINNA, I. DARRAULT–HARRIS (eds.) *Collection Actes: Formes de vie et modes d’existence durables*, Albi, CAMS/O, pp. I–XIX, <https://mediationsemiotiques.com/archives/9431> (consultato in marzo 2022).

USING TRANSFORMER NETWORKS AND TENSIVE SEMIOTICS TO IMPROVE SENTIMENT ANALYSIS ACCURACY IN TOURISM DIGITAL PLATFORMS⁽¹⁾

TULIO FERREIRA LEITE DA SILVA*

TITOLO IN ITALIANO: servirsi dei Transformer networks e della semiotica tensiva per migliorare l'accuratezza della sentiment analysis nelle piattaforme digitali legate al turismo

ABSTRACT: quella della *sentiment analysis* è una delle aree in cui il *natural language processing* trova maggiore applicazione, per via del suo interesse economico. Fin dai tempi in cui il passaparola era l'unico modo per condividere un'opinione, sapere cosa si dice di un marchio o di un prodotto è sempre stato molto importante nel processo decisionale delle aziende. Oggi, con l'avvento dei social media digitali come Tripadvisor, la condivisione delle opinioni è diventata molto più semplice. Tuttavia, elaborare questo volume di informazioni, noto come *Big Data*, non è un compito banale. Per questo motivo, si fornisce qui una presentazione critica dei modi in cui si conduce la *sentiment analysis* e si propone una nuova serie di categorie per classificare le posizioni delle persone, in sostituzione delle tradizionali "positiva", "negativa" e "neutra". A tal fine, si fa ricorso al cosiddetto punto di vista "generalista" della semiotica, proposto da Saussure e sviluppato da Hjelmslev, nonché ai progressi proposti in questo ambito dalla semiotica tensiva di Zilberberg. Al di là della critica, si presenterà infine una breve introduzione alle reti neurali artificiali in generale, e ai *Transformer networks* in particolare, con lo scopo di illustrare le evoluzioni del settore.

KEYWORDS: Text Mining; Tensive Semiotics; Zilberberg; Natural Language Processing; Deep Learning

(1) This research has received funding from São Paulo Research Foundation (FAPESP) — Grant number: 2020/15160-7.

* Universidade de São Paulo.

1. Introduction

In this Article, we intend to present the hypothesis, currently in verification, of the optimizing function of discursive semiotics for artificial intelligence. To this end, we use sentiment analysis as our scope, and present a general outline of automatic classification work. After the short introduction, we intend to discuss the categories *positive*, *negative* and *neutral*, commonly used in sentiment analysis, and critique them from the so called “generalistic” point of view (Saussure 2002; Hjelmslev 1969). Then, we propose a new system for the classification of restaurant reviews on Tripadvisor. Finally, we delve into *Transformer networks*, the state of the art in artificial intelligence methodology with regards to natural language processing (NLP).

2. Sentiment analysis

Sentiment analysis is the computational activity that looks to infer, from a given text, one of the following categories: *positive*, *negative*, *neutral*, and, in some cases, *mixed*. For this, according to the revision in Donega (2017), there are two broad lines of procedure: with dictionaries and ontologies, or with machine learning. Among the hypotheses presented in Silva (2021), the first one seems quite problematic when it comes to social networks. After all, the producers of texts in these platforms usually have a particular, and often peculiar, way of using language, which causes an exponential increase in the number of necessary rules for running the code effectively.

The discursive complexity of the internet, according to Barros (2015) is characterized mainly by its strong orality, in spite of its written form. Vowels are prolonged to emphasize certain situations (*reeeeeeally cool*), new words are created to express underrepresented phenomena, deviations from formal norms are extremely common, as are typing errors. Put together, these facts make dictionaries and ontologies impracticable to produce.

Nowadays, and possibly for a long time to come, classification activities are better resolved by a subgroup of AI tasks known as *machine*

learning. There are countless available methodologies and algorithms for such activities. However, according to Bengfort, Bilbro and Ojeda (2018), *neural networks* seem to be the most efficient methodology to handle the complexity of internet discourse: “neural networks are not only increasingly practical, they also promise nontrivial performance gains over traditional models. This is because unlike traditional models, which face performance plateaus even as more data become available, neural models continue to improve” (Bengfort, Bilbro and Ojeda 2018).

Inspired by biological models, deep learning algorithms utilize a computational architecture quite different from that of traditional code. Through the use of activation weights and functions, “intelligent” computational cells process only the data for which they were designed. Countless cells arrayed in deep layers allow different data characteristics to be grouped and processed in tandem and concurrently to other cells that group and process different characteristics.

Because of these *sui generis* characteristics, *neural networks* are capable of continually discovering patterns. Neologisms or deviations from standard language, for instance — which would cause many headaches to the dictionaries and ontologies programmers —, are readily perceived by *deep learning* algorithms. Understandably, they are considered the state of the art in NLP.

Despite this marvel of computation, sentiment analysis is regarded by many researchers in the field as one of the great challenges of artificial intelligence today. Goularas and Kamis (2019) indicate, through countless *neural network* configurations, that the maximum accuracy achieved by the state of the art is below 65%, which is unsatisfactory, and it means this computing technique is still an open field for development: “the accuracy achieved with the above techniques is high but still not satisfactory, thus making sentiment analysis an ongoing and open research subject. For this reason researchers try to develop new methods or improve the present ones” (Goularas and Kamis 2019).

Our hypothesis regarding the low performance connects it to the categories used for classification. In our view, measuring a text’s polarity with *positive*, *negative* and *neutral*, as obvious as it might be, causes serious problems, such as analytical ambiguity and high cost for annotated

corpora production. O'Hare et al. (2009) reported these problems a decade ago, when a project to produce an annotated corpus for reviewing the polarity of financial blogs demanded seven people to finish the task.

As sentiment annotation is a difficult task, and since domain knowledge of financial markets is necessary for annotating this corpus, it was important that our annotators were trained before undertaking this annotation task. The corpus was annotated by 7 people, 5 of these being computer science researchers from DCU, and 2 employees of our industrial partner. The training phase involved two rounds of pilot annotations consisting of 5 training documents each, followed by extensive discussions of these annotations, until a consensus annotation was reached. Following this, a set of guidelines for annotations was produced for the annotators. (O'Hare et al., 2009)

Despite being part of a computational field which seeks to monitor the subjectivity of text, using categories as subjective as *positive*, *negative* and *neutral* seems to further escalate an already exhausting task: “annotation of sentiment can be a relatively difficult challenge, as interpretation of sentiment is subject to a number of human factors such as domain expertise, the annotator’s private state and inferences the annotator has made into the text to be annotated” (O'Hare et al. 2009).

To further accentuate the difficulties of polarity recognition in text, it is worth mentioning that similar AI tasks that seek to identify shapes within images (a field known as computer vision) usually set off from an accuracy of 85% (Nara et al. 2019). That is 20% better performance than the best neural network algorithms presented by Goularas and Kamis (2019).

Another important matter being discussed is the analysis’ level of granularity, which can vary between document, sentence or aspect (Lv et al., 2021). On the level of document, there is a global sentiment perceived in the text as a unit. On the level of sentence, each one is taken as the smallest component and, first, classified as having an opinion or not, and then which polarity it possesses as a whole. Finally, on the level of aspect, the analyst strives for disambiguating complex sentences, where different components are identified as *positive* and *negative* simultaneously. For instance: “this restaurant has good food, but terrible

service” (Lv et al., 2021). The capacity for detection at aspect level is regarded as highly impacting for correct classification. When this level is ignored, the prediction error is usually close to 40%.

3. Generalistic point of view

Started by Saussure and formalized by Hjelmslev, a generalist point of view has for a long time been denouncing the hypostasis of language. For these thinkers, terms such as *positive*, *negative* and *neutral*, in spite of being concrete and somewhat unanimous, detract from the discussion more than they add to it. In his *Prolegomena*, the Danish master clearly denounces this phenomenon, which he believes is caused by the inconveniences of induction.

Experience alone is sufficient to demonstrate the obvious shortcomings of this method [inductive]. It inevitably leads to the abstraction of concepts which are then hypostatized as real. This realism (in the mediaeval sense of the word) fails to yield a useful basis of comparison, since the concepts thus obtained are not general and are therefore not generalizable beyond a single language in an individual stage. All our inherited terminology suffers from this unsuccessful realism. The class concepts of grammar that are obtained by induction, such as “genitive,” “perfect,” “subjunctive,” “passive,” etc., afford striking examples of this fact. None of them, as used until now, is susceptible of general definition: genitive, perfect, subjunctive, and passive are quite different things in one language, Latin for example, from what they are in another, say Greek. (Hjelmslev 1969)

The same criticism, however, was already offered by Saussure.

It will always be a theme of philosophical reflection that, for fifty years, linguistic science, born in Germany, developed in Germany, loved in Germany by countless individuals, has never had the fickleness to rise to the necessary degree of abstraction to master, on the one hand, what is done and, on the other hand, to what extent what is done has legitimacy and reason for being in the whole of science ⁽²⁾. (Saussure 2002)

(2) All excerpts in other languages were translated to English.

According to the Genevan master, in order to avoid such problems, we must take the following steps:

- a) describe all languages one can handle and their histories, which is equivalent to making the history of language families and reconstructing, as far as possible, the mother languages of each family.
- b) search for the forces that intervene permanently and universally in all languages, and draw up the general laws to which all particular phenomena of history can be reduced.
- c) demarcate and define itself. (Saussure 2002)

Something that seems to match the empirical principle proposed by Hjelmslev.

The description shall be free of contradiction (self-consistent), exhaustive, and as simple as possible. The requirement of freedom from contradiction takes precedence over the requirement of exhaustive description. The requirement of exhaustive description takes precedence over the requirement of simplicity. (Hjelmslev 1969)

Free of such concerns, the Genevan master punctuated that the linguistics of his time had no method and could not even know its object.

This [comparativist] school, though, that had the indisputable merit of opening a new and fertile field, did not come to constitute the true science of linguistics. It was never concerned with determining the nature of its object of study. Well, without this elementary operation, a science is unable to establish a method for itself. (Saussure 2002)

Contrary to the inductive method, using the semiotic apparatus in category building for artificial intelligence is justified due to its deductive principle. According to Meunier (2021), no matter the framework being used, semiotics (among other sciences) is privileged to be a theory-driven enquiry valid while a counterexample is not presented.

Saussure's and Hjelmslev's efforts to champion a scientifically oriented methodology notwithstanding, the continuing use of categories

like *positive*, *negative* and *neutral* for representing language phenomena leads to the realisation that the paradigmatic crisis of comparative grammarians extends to this day. To sweeten the bitter pill of criticism, we bring up Madaleine Arnold, who was, as early as the 1980's, pointing to semiotics as a sophisticated framework for category creation, and thus, a good tool for knowledge representation in Artificial Intelligence.

Semiotics has a global theory of meaning, which artificial intelligence currently lacks. Thus, in semiotics, the analysis of a text — be it verbal, iconic or other — benefits from a general economy which allows the pooling of certain knowledge and the distribution of others of a more specific nature, as well as feedback from specific analyses on the theoretical apparatus. (Arnold 1989)

By itself, the use of common-sense for representing linguistic phenomena appears epistemologically problematic. Despite this, if we wished to avoid this thorny, albeit necessary, discussion, applying semiotics as a tool for knowledge representation in AI would suffice to infer its optimizing function — since it is a framework built according to science. Moreover, according to Meunier (2021), the deductive apparatus can be successful because it uses a high quality method in identifying the features of a particular semiotic artifact,

As such, in order to present the framework developed by Zilberberg and known as tensive directions and segments (2006, 2012, 2019), in the next section we will demonstrate the applicability of the system in categorizing restaurant reviews.

4. Subjective quantifications

In Silva (2021), we deduce the whole formal apparatus of a restaurant review from the canonical narrative schema of Greimas. It is worth mentioning that from the dictionary definition of restaurant, which reads “a commercial establishment where meals are served to the public for a fee”, beyond the concept of utterance narrativization proposed by Barros (1987), it is possible to configure the producer of the discursive

subclass “restaurant review”. According to this apparatus, we propose characterizing the actor of the utterance as a syncretism between the actant roles subject of state and sanctioning addresser along with the thematic role *Tripadvisor user*.

For this paper, though, we must set aside all these frameworks and focus solely on the model called “Tensive directions” (Zilberberg 2006, 2012, 2019), a diagram adaptation (figure 1) will be used to illustrate this model. Applied to texts where a review or an appreciation are manifested, the diagram displays *subjective quantifications*, a term coined by the Brazilian semiotician Luiz Tatit in 2011.

Both the feeling of lacking and the impression of excess presuppose an evaluator’s intermediation, the so-called judging addresser, who, within an axiological framework, establishes their measuring parameters, that are arbitrary from the point of view of scientific exactness, but easily recognizable as common-sense in a given sociocultural group. (Tatit 2011)

With *plus* and *minus* particles, which Zilberberg (2006, 2012, 2019) calls *tensive syllables*, he provides us with an apparatus capable of measuring a verifiable gradation in both the objective and the subjective dimensions. For example, although it is hard to measure the amount of love in a letter, it is possible to quantify, with a comparison, rather the sentiment appears to a higher or lesser degree. In order to exemplify the use of Zilberberg’s system and metalanguage, we have selected a few reviews of Alex Atala’s restaurant D.O.M.

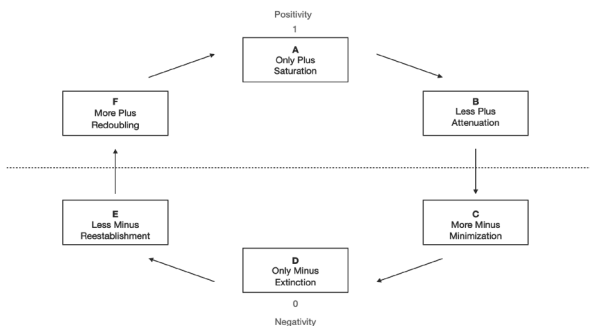


Figure 1. Diagram of Zilberberg’s tensive directions (adapted from Tatit 2019).

4.1. Redoubling

Seemingly the most abundant category on Tripadvisor, it encompasses texts built from a distribution and gradual enumeration of *pluses*.

Without the shadow of a doubt, the best gastronomical experience I've ever had, with the privilege of having distinctively Brazilian ingredients in the dishes. There's no doubt that the dear and very thoughtful Alex Atala is a genius! (Silva 2021)

It is easy to perceive that, in this case, the *shifters* focus on the relation between addresser (user) and addressee (restaurant): “[I]’ve ever had” and “[the restaurant offered] the best gastronomical experience”. In the example above, the redoubling is recursive, since the *pluses* continue to be introduced: “distinctively Brazilian ingredients”, the thoughtfulness of the chef, etc.

The category of redoubling represents quite well the traditional *positive* (from the classification that also includes *negative* and *neutral*). In it, the reviews where positivity is exclusively present are gathered. Be it at the level of document, sentence or aspect, redoubling represents the *plus* in constant progression. In this category, there is no room for negativity.

4.2. Saturation

Still in a *positive* direction, this category marks a textual disruption. In the following excerpt, it is easy to see the addresser–addressee relation is broken in favour of another, where the addresser shifts out an observer, a subject marked by the pronoun “you”, of whom we know very little.

If you go to DOM, you opted for a gastronomical experience, not a mere dinner. The restaurant is quite expensive, but it's worth every penny. I recommend, for your experience to be the best one possible, that you opt for the Maximus menu with harmonization. It's 12 dishes in total, and 11 beverages. Each dish and each beverage has a history and relation with Brazil and small farmers. With each dish, we are surprised by the creativity, delicacy and combination of flavours. You're gonna eat ants... and you're gonna love it. (Silva 2021)

This is written as if the utterance actor had entered a conjunction so deep with the object meal, in an ecstasy so great with the gastronomical experience, that they are talking to themselves. In this case of plenitude, we clearly see the conjunction relation is overflowed, configuring a fusion.

In semiotics, saturation is marked by a rupture with expectations. In the case of a simple restaurant review, for instance, in linguistic terms, one always expects an “I” as the reviewer and a “you” as the restaurant. Enunciative strategies may try to hide the pronouns, aiming for “objectivity”. However, if an utterance exists, it is because an utterer exists and, as much as “you” becomes “it”, the game between reviewer and restaurant remains. For example, “the restaurant is really good” and “I thought your service was great” characterize well two grammatical forms of reviewing.

However, in the presented excerpt, one can notice a break with expectations. In the review, “you” are no longer the restaurant, but a different person entirely, whom Semiotics conventionally names *observer*. In this case, the review becomes a recommendation instead. Despite the possible linguistic preciousness, the category of saturation can serve to classify some exceptions to traditional reviews. For instance, given the great success of reality cooking shows, it is not hard to find reviews in Tripadvisor that are about some event unrelated to the restaurant, and not the gastronomical experience. These are users that intend only to compliment the chef due to some political position, or even with picturesque characteristics, such as being very handsome, etc.

4.3. *Attenuation*

So far, we have only mentioned cases of full positivity. In this category, though, for the first time we come across a *minus*: “I really liked the tasting dinner menu, but I was expecting more because of the Michelin star. The service is great” (Silva 2021). In the attenuation, we are close to a certain degree of negativity. However, the arrival of a single *minus* is not enough to move this utterance from the field of *positive* reviews.

The category of attenuation, thus, handles the complexity in cases where a sentence may not be characterized as either *positive* or *negative*,

since both polarities are present within it. As pointed out by Lv et al. (2021), much of the difficulty in obtaining a sentiment analysis model derives from the non-differentiation of this category — which usually confuses the natural language processing algorithms.

The sentence “I really liked the dinner tasting menu, but I expected much more, given the Michelin star” clearly cannot be characterized as purely *positive*, but, at the level of document (the entire review), positivity is much more present. Therefore, the category of attenuation, something like *positive, but*, is characterized by the presence of negativity in a document where positivity is the main direction.

4.4. *Minimization*

Differently from attenuation, minimization is truly the first purely negative representative. In this category, as we see in the following excerpt, there may even be some *pluses*, but the *minuses* are dominant, configuring completely the negative direction.

The house’s 20-year-old tasting menu starts out lukewarm and reaches its peak with the pirarucu with paçoca... could’ve stuck with the cassava variations. The disappointment came when we received the dish list at the end, and realised our dish was incomplete — an amateur’s mistake that one wouldn’t expect from a house as renowned as DOM. When we called to report what had occurred, we were met with disdain. If any doubts remained about the 50 rank regression in the top 50 list, they’ve been cleared up. (Silva, 2021)

The conjunction relation between subject and object remains present and very clear. Still, the utterance actor, motivated by his axiological framework, denounces a breach of contract: “our dish was incomplete”. The restaurant’s unsatisfactory performance does not end there. Seeking compensation, the utterance actor was met with disdain. In this text, we notice that the *minuses* are increasingly presented in a syntagma clearly marked by a tensive decadence.

4.5. *Extinction*

As in *saturation*, in this category, we observe a rupture with the traditional actantial logic. Worse than a disjunction, in the following excerpt, we see the transformation of the object into an abject⁽³⁾.

Me and my wife tried to book a reservation at DOM and the e-mail that reached us said there weren't any tables available for our date and time. Then, I sent a reply saying I had no interest in the other dates they offered, since I wanted to go on her birthday (January 30th). Now I've just checked my credit card and there's a charge of 2 hundred reais for not showing up! I went to check my e-mails and, out of nowhere, they said a table had opened up! Even after I sent an e-mail giving up any dates! That's dirty! Now I have to spend 2 hundred bucks for not going, while not knowing there was a table for me! Because I had already given up! I didn't even check my e-mails again since I'm in town for vacation! Dom! Go f... yourselves! (Silva 2021)

From the highest of goals, a celebration in style of his wife's birthday, the conjunction with the object becomes something undesirable. It is clear in the text the object's metamorphosis into abject. We notice the utterance actor develops little by little a veritable repulsion by the restaurant, a path that ends with (and is figurativized by) the not fully realized cursing: "Dom. Go f... yourself!"

Also among the cases worth mentioning, the boycott is the complete extinction of the narrative program "restaurant review". In it, according to Greimas' propositions, it is clear that the sanction ceases being a base program (sanctioning the junction with the meal), to become a use program within a path that belongs only to the utterance actor.

We have the civic and humanitarian duty to eliminate this restaurant from the map. Senseless owner. Statements that show the company's values and mission. This is my opinion. (Silva, 2021)

(3) Cf. Zilberberg C. (1988) *Raison et poétique du sens*, Presses Universitaires de France, Paris.

The text, titled *Genocidal owner*, clearly makes no mention of a meal. It is an attempt to boycott a Brazilian hamburger chain, whose owner had made statements minimizing the Covid-19 pandemic and railing against lockdown policies.

4.6. Reestablishment

Unfortunately, we did not find examples of this category so far. However, since systems are predictability models (presenting the possible, as opposed to the realised), we can infer examples of this category: “After the disaster in our last visit, we at least received a voucher as an apology. We might return”. Here, as with attenuation (though in the opposite direction), we notice an attempt to leave the zone of negativity towards positivity, which is still unrealized. That is to say the chain of *minuses* remains predominant, but *pluses* begin to appear.

To finish this section, it is worth mentioning that the length of the text does not matter. As shown by Tatit (2011, 2019), the elasticity of language fits into the mechanism developed by Claude Zilberberg (2006, 2012, 2019). Moreover, the mechanism is capable of limiting the exact place of the reacting subject, like in a photograph, crystalizing the publishing moment. However, sentiment analysis models that take other levels (than the whole text) into consideration can also make use of the concepts of *plus* and *minus* to classify their corpora. As the entire text is a syntagma, it allows the discovery of *plus* and *minus* figures dressed as sentences or phrases.

5. Transformer Networks

Currently, a new network is responsible for a veritable revolution. Named *Transformer* by the *Google Brain* and *Google Research* teams, this new architecture takes a single component from recurrent neural networks and places it in the center of the model’s functioning. Presented in the article *Attention is all you need* (Vaswani et al. 2017), the model is entirely based in the structure of *neural attention*.

According to Rothman (2021) this new architecture is revolutionary and disruptive because it changes how natural language

processing and artificial intelligence itself are conceptualized. He claims the Transformer network is not an evolution, because it breaks with previous models (recurrent and convolutional networks) and can surpass human intelligence within the next few years.

The *Transformer network* has already consolidated itself as the state of the art in most NLP tasks (Rothman 2021), with faster training and better performance rates than previous architectures.

The traditional model was developed to translate between different languages and it is composed by an encoder and a decoder. Verbal input goes through an embedding process⁽⁴⁾ and the addition of noise that marks the position of words within sentences (positional encoding). Output embedding is the stage where the model receives the sentences in the target language, to which the input, in the original language, must be translated.

Each module (encoder and decoder) was multiplied sixfold in the original article. Layers of multi-head attention process countless mathematical calculations (scalar product, among others) in parallel. The skip connection technique is present in all stages, which avoids the explosion or disappearance of the gradient. Finally, a softmax layer distributes the probabilities.

Currently, traditional architecture is heavily modified for several activities (Rothman 2021). With only the decoder module, for instance, Open.AI developed the GPT models, responsible for the state of the art in text generation. At the same time, by evolving the encoder, Google has created BERT (Bidirectional Encoder Representations from Transformers), which currently surpasses countless architectures in tasks such as summarizing and sentiment analysis, among others.

According to Lopes et. al. (2021), BERT is an artificial intelligence methodology that supervises itself, generating language models. From a great amount of pure verbal data, the networks are trained without the need for an annotated corpus. Once the model with the representation of that language is obtained, it can be used for specific PLN workloads, with the need of no more than some quick training in which parts of the model adapt to the task it must solve.

(4) As computers cannot process letters, word embedding — a class of techniques that maps words from some corpus to vectors of real numbers — is used.

This fine tuning of networks is part of the “transfer learning” paradigm, in which a previous neural network’s initial layers are frozen and reused, bypassing the need for training in generic characteristics identification in images (such as curves, contours and horizontality or verticality). Thus, only the last layers in pre-trained models need to undergo new training, severely reducing computing costs.

BERT was initially modeled in the English and multilanguage versions. But the revolution and interest it awakened in the field was such that countless other models (in several languages) are constantly being launched. In 2020, a Brazilian project from Unicamp presented BERTimbau, a network trained in Brazilian Portuguese (Souza et al. 2020; Souza, 2020).

BERTimbau represents a Brazilian advancement in the utilization of deep learning in natural language processing. Pre-trained with Portuguese Wikipedia articles, the model has a rich vocabulary from which it can model the Portuguese language. When validated, BERTimbau re-established the state of the art in Brazilian Portuguese NLP tasks such as text similarity, named entities, among others.

6. Conclusion

Our goal with this paper was to show semiotics’ heuristics as a powerful tool for natural language processing in artificial intelligence, and more specifically, sentiment analysis with artificial neural networks. If our hypotheses are correct, therefore, using a corpus annotated with Zilberberg’s proposition, we believe it is possible to obtain a significant jump in sentiment analysis performance and granularity. This is because for the first time, sentiment analysis would be done with a more encompassing, less ambiguous model than the current one. To this end, after the first proof of concept presented in Silva (2021), we are using algorithms from Transformer networks to verify our solution’s degree of accuracy applied to sentiment analysis.

References

- ARNOLD M. (1989), *La sémiotique: un instrument pour la représentation des connaissances en intelligence artificielle*, “Études littéraires”, vol. 21, n. 3, Paris.
- BARROS D. (1987), *Problemas de enunciação*, “Revista Cruzeiro Semiótico”, n.6, Portugal.
- (2015), *A complexidade discursiva na internet*, “CASA: Cadernos de Semiótica Aplicada”, v.13, n.2, Brazil.
- BENGFORT B., BILBRO R., OJEDA T. (2018), *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*, O’Reilly Media, USA.
- DONEGA T. (2017), *Os efeitos das revisões críticas online sobre o mercado cinematográfico americano*, “Dissertação de Mestrado (Programa de Pós-Graduação em Sistemas de Informação)”, EACH-USP, Brazil.
- GOULARAS D., (2019) S., *Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data*, “2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)”, Turkey.
- HJELMSLEV L. (1969), *Prolegomena to a theory of language*, The University of Wisconsin Press, USA.
- LOPES E., CORREA U., FREITAS L. (2021), *Exploring BERT for Aspect Extraction in Portuguese Language*, “The International FLAIRS Conference Proceedings”, n. 34, USA.
- LV *et al.* (2021) *Aspect-level sentiment analysis using context and aspect memory network*, “Neurocomputing”, v. 428, International.
- MEUNIER J-G (2021), *Computational semiotics*, Bloomsbury Academic, USA
- NARA, M., P. PADALA AND B. KINNAL (2019) *Performance Evaluation of Deep Learning frameworks on Computer Vision problems*, “Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)”, India.
- O’HARE N. *et al.* (2009), *Topic-Dependent Sentiment Analysis of Financial Blogs*, “TSA ‘09: Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion”, USA.
- ROTHMAN D. (2021), *Transformers for Natural Language Processing*, Packt Publishing, USA.

- SAUSSURE F. (2002), *Écrits de linguistique générale*, Éditions Gallimard, France
- SILVA, T. (2021) *Inteligência semiótica e TripAdvisor: deep learning e quantificações subjetivas na análise dos discursos de sanção de restaurante*, “Dissertação de Mestrado (Programa de Pós-Graduação em Semiótica e Linguística Geral)”, FFLCH-USP, Brazil.
- SOUZA F. (2020), *BERTimbau: pretrained BERT models for Brazilian Portuguese*, “Dissertação de Mestrado (Programa de Pós-Graduação em Engenharia da Computação)” Unicamp, Brazil.
- SOUZA F., NOGUEIRA R., LOTUFO R. (2020), *BERTimbau: Pretrained BERT Models for Brazilian Portuguese*, “Intelligent Systems. BRACIS 2020. Lecture Notes in Computer Science”, vol. 12319, Springer, Cham.
- TATIT L. (2011), *Quantificações subjetivas: crônicas e críticas*, “Cadernos de Letras da UFF — Dossiê: Linguagens em diálogo”, n. 42, Brazil.
- . (2019), *Passos da semiótica tensiva*, Ateliê Editorial, Brazil.
- VASWANI A. et al. (2017), *Attention is all you need*, “Advances in neural information processing systems”, USA.
- ZILBERBERG C. (2006), *Éléments de grammaire tensive*, Pulim, France.
- (2012), *La structure tensive*, Presses Universitaires de Liège, Belgium.
- (2019), *Horizontes de la hipótesis tensiva*, Fondo Editorial/ UniLim, Peru.

HOW TO BUILD A CHATBOT. A SEMIOTIC AND LINGUISTIC APPROACH

ANNA DALL'ACQUA, FEDERICO BELLENTANI*

TITOLO IN ITALIANO: *Come creare un chatbot: Un approccio semiotico e linguistico*

ABSTRACT: chatbots are Artificial Intelligences able to simulate a written or vocal conversation in natural languages through software, applications or websites. Their aim is to provide users with the information they need through a faster and simpler user experience. Many businesses and organisations use chatbots as tools to improve user experience, provide 24/7 customer service, simplify online shopping and payments, manage reservations and so on.

Semiotic research has recently analysed chatbots. Its aim has mainly been to advance the understanding of how to make the conversation of chatbots “more human” as possible in order to avoid disruption and frustration in users. Semiotic analysis is accurately used to respond to this challenge, focusing on how humans make meanings and interpret them, against a machine-centred approach. However, it makes theoretical generalisation from analysis of existing chatbots in specific domains, e.g. healthcare. The domain in which a chatbot operates is important, so is the cultural context, that is notably absent in previous semiotic analysis of chatbots.

This paper presents a semiotic and linguistic approach to chatbot design that focuses on three dimensions: a) the user/chatbot goal; b) the target user, her context and interpretations; c) the chatbot personas. This approach can prove useful to identify how users make and interpret meanings during conversations with chatbots in a given social and cultural context.

KEY-WORDS: Chatbot, Conversational AI, Human-Computer Interaction, Semiotics, Linguistics

* Future Education Modena; Università degli Studi di Torino e Injenia.

1. Introduction⁽¹⁾

Chatbots are artificial intelligences able to simulate a written or vocal conversation in natural languages through software, applications or websites. Their aim is to provide users with the information they need through a fast and simple user experience, without the need to use the device's memory to download or update them. Many businesses and organisations have started to consider chatbots as tools to improve the customer experience: by 2025, the chatbot market size is projected to grow to \$10.4 billion (*Market Data Forecast 2020*).

There is a terminological confusion in conversational AI applications in the academic, public and business discourse: “chatbot”, “virtual assistant” and “conversational agent” are often used as synonyms (Dale 2016). This paper uses the terms “chatbot” or “conversational agent” to generally refer to conversational AIs that interact with users through natural language in a conversational manner (McTear 2020). Specifically, “chatbot” is the oldest term referring to a conversational system designed for specific purposes such as improving the user experience of company websites or e-commerce platforms, providing 24/7 customer service, simplifying online shopping and payments, managing reservations, generating and qualifying marketing leads. Gnewuch *et al.* (2017) group chatbots in the industrial sector into five categories: generic customer-service, help-desk, website navigation support, shop assistant and technical support. According to them, customer service is the most common application of chatbots in business. They are employed also in healthcare (Jovanovic *et al.* 2020) and education (Winkler & Söllner 2018), especially in languages teaching (Fryer *et al.* 2020).

Some semiotic research has recently analysed chatbots. Valtolina, Barricelli & Di Gaetano (2020) studied the communicability of chatbot in comparison to traditional graphical user interfaces. They demonstrated that chatbots can help users with poor technological attitudes to solve simple tasks by analysing cases in healthcare and smart home domains. Valério *et al.* (2017: 28) used a semiotic inspection method

(1) This paper has been written in equal parts by the two authors. For formal attribution, the parts 2 and 3.1 can be attributed to Anna Dall'Acqua and the parts 1, 3.2 and 4 to Federico Bellentani.

to analyse the communicative strategies of chatbots, i.e. “a qualitative evaluation method based on Semiotic Engineering that assesses systems’ communicability”. The authors identified strategies for chatbots to present their features to users and thus being more effective. Klimczak, Wirsching & Graben (2020) claimed that current speech assistive devices can be frustrating because they still lack semantic analysis capabilities. To overcome this shortcoming, they demonstrated that machines only need to learn human meanings by a trial-and-error process, without understanding the meanings of single words or implicit context knowledge. The aim of the semiotic analysis of chatbots is to understand how to make the conversation of chatbots “more human” as possible in order to avoid disruption and frustration in users. Semiotic analysis is accurately used to respond to this challenge, focusing on how humans make meanings and interpret them, against a machine-centred approach. However, semiotic research on chatbots makes theoretical generalisation from analysis of existing chatbots in specific domains, e.g. healthcare. The domain in which a chatbot operates is important, so is the cultural context, that is notably absent in previous semiotic analysis of chatbots.

This paper proposes a semiotic and linguistic approach to the phase of chatbot design that takes place before the actual implementation of the solution, whose aim is to identify the project’s requirements and design the most appropriate human-computer interaction. To do so, section 2 first sets the context by providing an historical and cultural overview of chatbot and conversational AI. Section 3 introduces the framework focusing on the three dimensions of chatbots: a) the user/chatbot goal; b) the target user, context and interpretations; c) the chatbot personas.

2. Historical and cultural overview of chatbot and Conversational AI

2.1. Key terms in Conversational AI

The term “chatbot” comes from “chatterbot”, coined by the computer scientist Michael Mauldin (1994) to describe conversational softwares.

He named “chatterbot” his game character for a multiuser game whose main function was to chat. The term “chatterbot” is still used in the literature (Luger & Sellen 2016), together with its variants “chatbot” (Panesar 2019), “chat bot” (Wuenderlich & Paluch 2017), “chat-bot” (Bala *et al.* 2017), “chatter bot” (Chakrabarti & Luger 2015). All these terms refer to “computer programs that interact with users using natural languages” (Shawar & Atwell 2007: 29) or to systems designed for extended unstructured conversations mainly for entertainment purposes (Jurafsky & Martin 2019). The website *chatbots.org/synonyms/* shows the proliferation of terms used to name these technologies: beside multiple spelling variants of chatbots, there are several terms related to the domain of dialogue and assistance.

The primary modality of interaction used by these technologies is traditionally text-based, but chatbots used through vocal commands are becoming increasingly more common: in this case, the term “voicebot” is used to highlight this feature and they are often associated with avatars or talking heads (Ciechanowski *et al.* 2019). Moreover, studies such as Dasgupta (2018) and Pearl (2016) support the opportunity of a multimodal modality of interaction that involves multiple channels. Radzwill & Benton (2017) identify chatbots as conversational agents that mimic human interactions without any humanoid robots, which are typical elements of the embodied conversational agents.

Chatbots belong to the wider category of dialogue system, a generic term referring to software that communicates with users in natural language in the form of text, speech or both. Jurafsky & Martin (2019) consider dialogue systems and conversational agents as synonyms and categorise them according to their functions of completing tasks. There are several classifications of dialogue systems based on their functions and goals in the literature (Hussain *et al.* 2019). The task-oriented conversational systems include virtual assistants whose interactions are usually short, including commands and notification tasks across a broad range of contents (Jacques *et al.* 2019). No-task-oriented conversational agents are designed for extended and unstructured conversations without any specific task to achieve (Cercas Curry *et al.* 2017). Mathur & Sing (2018) distinguish chatbot in two groups: “goal-driven” and “non-goal-driven”, associable to “task-oriented”

and “no task-oriented” respectively. Task-oriented chatbots have become more popular thanks to social media platforms such as Facebook, which started to use them to help users to achieve specific tasks, such as scheduling a meeting or booking a hotel (Hussain *et al.* 2019).

Personal assistants such as *Google Assistant*, *Amazon Alexa*, *Apple Siri* and *Microsoft Cortana* are often compared with chatbots (Chaves & Gerosa 2019). They are voice-activated assistants that can respond to a much larger number of commands than the task-oriented chatbots. They are always connected to the Internet and each interaction is analysed by a central computer system that matches the voice’s user commands with the assistants’ response (Hoy 2018). McTear *et al.* (2016) provide a list of terms used to mean these objects, which includes “personal assistant”, “virtual personal assistant”, “intelligent personal assistant”, “digital personal assistant”, “mobile assistant” and “voice assistant”.

In the taxonomy proposed by Følstad *et al.* (2019), the term “chatbot” refers to both text-based and voice-based conversational agents and encompasses personal assistants such as the Google Assistant as well. The classification is based on two high-level dimensions: locus of control and duration of interaction. The first dimension classifies agents according to their role in leading the conversation (chatbot-driven dialogue and user-driven dialogue); the second one according to the duration of relation (short-term relation and long-term relation). The locus of control comes from the taxonomy proposed by Zue & Glass (2000) in “user-initiative”, “system-initiative” and “mixed-initiative” interfaces. Hussain *et al.* (2019) summarises the broad classification following four criteria: interaction mode, goal, implementation technique and domain.

The fact that there is not a single definition for chatbot and Conversational AI technologies is meaningful in itself: this lack demonstrates inconsistencies in the evaluation (Braun & Matthes 2019) and in the studies on the interaction design techniques for their implementation. Since there are terminological inconsistencies and overlappings, it is difficult to identify a coherent design methodology suitable for the various modalities of interaction that chatbot and Conversational AI offer (Fadhil & Schiavo 2019).

2.2. *An historical overview of chatbots*

The first chatbots were based on pattern matching and simple template-based mechanisms with rudimentary conversational abilities (Weizenbaum 1966; Colby *et al.* 1971; Wallace 2003). Traditional building processes were rule-based, with the system's behaviour determined by a predefined set of handcrafted rules and therefore predictable. They used rule-based grammars for the Natural Language Understanding (NLU) component to interpret utterances produced by users.

The 1950 article *Computing Machinery and Intelligence* by Alan Turing can be considered as the origin of the study of Human-Computer Interaction. Here Turing was asking the question "Can the machine think?" and proposed to evaluate machine thinking by examining their capacity of producing and understanding natural language. Even though the creation of a machine able to talk was not a priority, Turing laid the foundations for the first ancestors of chatbots.

The first actual chatbot traces back to 1966, when the computer scientist Joseph Weizenbaum implemented ELIZA, known also with the name of DOCTOR. The conversation of this chatbot mainly consisted of questions, whatever was the users' inputs. The idea was to impersonate a Rogerian therapist. The psychological attitude characterised also the chatbot PARRY (Colby *et al.* 1971), that simulates the behaviour of a paranoid schizophrenic patient whose responses were usually based on conceptualizations and misunderstandings. Both these chatbots were realised through a simple pattern matching technique and a rule-based approach (Jurafsky & Martin 2019). An analogous rule-based approach was the one underlying A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*), a pattern-matching chatbot labelled through the markup language AIML derived from the metalanguage XML (Wallace 2003).

Despite these predecessors, only recently chatbots and conversational agents have become popular: Dale (2016: 811) identifies chatbots as "one of the most breakthrough technologies of 2016", while Tsvetkova *et al.* (2017: 1) speaks about an "ecosystem of bots". Technologies such as *Google Home* or *Alexa* are taking over households, making

conversational technology become a part of everyday life. However, despite the important achievement in Natural Language Processing and the potential of conversational technologies, we are still far from creating a machine capable of truly understanding natural language and holding a real human-like conversation (Wu *et al.* 2019).

Today most conversational agents base their NLU component on machine learning algorithms. Machine learning led to the spread of conversational technologies and platforms (Dale 2016). Today, the largest tech industries have their own platforms, like *Dialogflow*, *IBM Watson* or *RASA*, but the data for the initial training⁽²⁾ are still private (Braun *et al.* 2017). The architecture of these systems is like the modular rule-based models, but the core modules are modelled probabilistically and learned from data (McTear 2020). The most recent research in conversational agents architecture focuses on end-to-end neural dialog systems, where an input utterance is mapped to an output response and therefore without requiring any processing by the modules of the traditional architecture (McTear 2020).

The commercial tools for building conversational agents share a statistically data-driven NLU core model, based on the Intent-Entity-Context-Response (IECR) paradigm (Moore & Arar 2019). This can extract useful information from naturally unstructured natural language inputs and to structure them into *intents* and *entities*. An “intent” is “a mapping between what the user says and what action should be taken by the chatbot” (Adamopoulou & Moussiades 2020: 377); “entities” are tools that do not deal with the entire utterance but with the extraction of details or parameter values inside the sentence itself. The “context” is a feature that helps to structure the conversation and the interplay between different dialogue states that need more than one single conversational turn to be developed. “Response” is the response that appears during the conversation. The response generation in the mentioned platforms is *retrieval-based* (Wu *et al.* 2017), i.e. that the system retrieves some response candidates from an index and analyses available information using APIs, before matching the input with

(2) These platforms adopt a supervised machine learning model and are trained on both private data provided by the proprietary industries and on input data specific for the single project selected, uploaded and annotated by the developer working on that specific project (Braun *et al.* 2017).

the corresponding output. Current research is focusing on Natural Language Generation algorithms using advanced deep learning algorithms, i.e. the generation of answers based on previous users' inputs; however, there are still difficulties in building and training them (Adamopoulou & Moussiades 2020).

3. A semiotic and linguistic approach to build chatbots

3.1. Conceptual frameworks to investigate conversations

A conversation is an exchange of ideas by spoken words. It is the tangible expression of our mutual understanding and values, which is expressed through language. Language, in turn, is shaped and defined by conversations. Conversations between humans are often seen as social and transactional talks (Clark *et al.* 2019). Social talks are made to develop, build and maintain relationships. Transactional interactions are aimed to achieve a specific goal. Typically, their primary goal is the effective transferring of information from one speaker to another (Brown & Yule 1983). Social and transactional talks are two categories that often overlap, depending on the context and the participants (Cheepen 1988). Schlegoff (1999) identifies the ordinary conversation as the most general type of dialogic interaction, from which other types of conversations are derived and adapted for specific purposes. Moore & Arar (2019) define the ordinary conversation as the most flexible type, covering a vast number of intentions. In ordinary conversation, there is a balance between the speakers, while in other types of conversation there can be asymmetry or subordination (Linell & Luckmann 1991).

Linguistics has traditionally explored conversational and dialogical exchanges. Functional linguistics has focused on the interactional and contextual dimension of language. Dialogical interactions are also at the core of pragmatics, which is known to be the field of linguistics involved in the study of language in use and their users in interaction (Bublitz & Norrick 2011). One of the most relevant frameworks of analysis in pragmatics concerning the structure of conversation is Conversation Analysis (Levinson 1983). The origins of this approach

can be traced back in the late 1960s, when the sociologists Harvey Sacks, Gail Jefferson and Emanuel Schegloff observed that conversations were not as chaotic as they could seem, rather they were based on regular internal structures. They performed an inductive data-driven analysis through collected video- or audio-recorded conversations to find recurring patterns of interaction. This analysis showed that each type of conversation depends on the same “conversational machinery” (Sacks 1984: 26), but the activities and settings in which they occur contain distinctive patterns and slight adaptations (Sacks *et al.* 1974; Schegloff 2007).

Conversation Analysis can be useful to identify structural elements of conversations in spontaneous human interaction that can be applied also to Human-Computer Interactions. These categories are: turn-taking (Sacks *et al.* 1974) sequential organisation in expandable adjacency pairs sequences (Schegloff 2007; Schegloff 1968; Schegloff & Sacks 1973), storytelling sequences (Jefferson 1978), repair practices (Schegloff *et al.* 1977), recipient design (Sacks & Schegloff 2007; Sacks *et al.* 1974) and minimization (Sacks & Schegloff 2007). The variety of approaches to verbal interactions analysis increased as Conversation Analysis crossed the borders of the USA and became progressively implanted in other countries, such as in France, where the methods of the conversation analysis were adopted mainly by linguists with a strong cultural background in discourse analysis (Kerbrat-Orecchioni 1997).

If Conversation Analysis positions itself on a structural and empirical point of view, dialogic interactions can be treated on the level of meaning and abstract principles (Grice 1975; Searle 1969). One of the key attributes of conversation between humans is indeed establishing common ground and mutual understanding between the speakers (Clark 1996), i.e. having the knowledge of what the other person is like in terms of personality, in order to contextualise his/her utterances. This attitude comes spontaneously: according to the Cooperative Principle (Grice 1975), humans naturally expect cooperation between the participants in a conversation. This is relevant in relation to conversational agents because humans tend to cooperate in the conversation with the chatbot as they would to another person (Reeves & Nass 1996), even though empirical studies demonstrate that final users do

not leave unnoticed they are actually conversing with chatbots (Hill *et al.* 2015).

These traditional linguistics approaches to conversations lay the foundation of multiple methodological texts (for example, Cohen *et al.* 2004; Pearl 2016) and of singular experimental studies (for example, Jacquet *et al.* 2018). However, the cultural and semiotic dimensions are neglected areas in this field. The next section makes up for this lack by providing a semiotic approach for chatbot design.

3.2. *Semiotic elements for chatbot design*

The linguistic research methodologies and perspectives presented in the previous section are useful to identify structural elements of human conversations and thus are applicable to human–computer ones. Here we present an alternative approach based on semiotics to implement culture and semiotic aspects into the chatbot design. Specifically, this approach focuses on the three dimensions that are to be considered before the actual conversation design and implementation of the chatbot: a) the user/chatbot goal; b) the target user, her context and interpretations; c) the chatbot personas.

a. Defining the user/chatbot goal: the actantial model by Greimas (1966: 174–185 and 192–212) can be useful to theoretically analyse potential user goals. In this model, goals are broken down into six components named actants: subject, object, helper, opponent, sender and receiver. The subject–object relationship establishes the axis of desire: the subject is directed toward a desired object that, in the case of chatbot, is the goal the user strives to achieve. The relationship between subject and object defines a junction, which becomes a conjunction when the subject connects with the object, i.e. when users get the information they need. Helpers may assist the subject in reaching the goal, while opponents hinder the same. The sender requests the junction between subject and object, while the receiver is the element for which the action occurs. All actants are analytical components, not ontological entities. Different actors can thus assume different actantial roles: for example, small talk or irony in chatbot can assume the role of helper for the user that is predisposed to them, while they can be opponents for the user that is eager to get information as fast as possible.

Actantial analysis can be a tool used to analyse the action that takes place in the conversation between users and chatbot and the functional relations between users and the information they need. As such, it can better explain the reasons behind the subject's informative needs and identify the elements that help or hinder the goal. The actantial model and narrative analysis may help to define helpers that engage with the users, suggest interests and provide content to let the users better recognise their needs.

Once the users' goals are clear, the first practical step is to align them with the scope of the chatbot. A successful chatbot design exhaustively covers the main users' needs, but only according to a series of defined paths. Chatbots can thus take care of simple tasks, but their conversation has to be articulated to effectively interact with users. The first questions to answer before the implementation are: what is the scope of the chatbot? What it can and cannot do? Answering those questions is crucial before starting the chatbot project to understand if the chatbot goal is feasible.

Narrative semiotics of Greimas (1970, 1983) can improve the understandings of the following dimensions of the users engaged in the conversation with the chatbot:

- the cognitive dimension refers to the knowledge users have, that affects how users engage with the chatbot, talk with it, evaluate new information;
- the axiological dimension considers users' personal opinions and evaluations. The question is whether users have positive or negative attitudes towards the conversation and the new knowledge gathered from it;
- the emotional dimension identifies which kinds of emotions and feelings users have and develop during the conversation;
- the pragmatic dimension concerns how users interact within the chatbot and how they act according to the knowledge gathered through the conversation.

b. Defining the target users, her context and interpretations: the model of textual interpretation by Umberto Eco (1990) — the “Model Reader” — analyses the complex interaction between authors, readers

and texts. He suggested that textual interpretation lies in an intermediate point between the authors' and the readers' intentions. Eco (1990: 50–51) dubbed this intermediate point “intention of the text” or *intentio operis*, that interacts with the *intentio auctoris* and the *intentio lectoris* — that are the intention of the author and the intention of the reader respectively. Envisioning the intention of the text has overcome the idea that appropriate interpretations occur only when readers follow the intentions of authors. In the wake of these proposals, semiotic analysis has begun to include interpretations deviating from the intentions of the authors. However, texts necessarily impose certain constraints on interpretation and make certain reading more desirable than others: textual strategies are available to authors to entice readers along a specific interpretation. Eco grouped these textual strategies under the terms “Model Reader” (Eco 1979: 7–11). According to this model, empirical authors write texts making assumptions about the readership's social background, education, cultural traits, tastes and needs. Hence, empirical authors foresee and simultaneously construct their readership, emphasising certain interpretations while concealing others (Eco 1979: 7–11). Although authors seek to control interpretations, texts do not function as mere “communicative apparatuses” to directly imprint meanings to readers (Eco 1984: 25). Yet, texts are aesthetic productions that inevitably leave something unexplained. As such, texts became the place where authors and readers continuously negotiate their interpretations: while authors empirically seek to control readers' interpretations, readers interpret texts in line with their knowledge, experience and needs. Hence, a complex interaction between authors, readers and texts themselves underpin textual interpretation.

This model can be applied to conversation design: chatbot designers strive to entice users along interpretations that conform to their intentions, anticipating a set of interpretations and discomfoting others. “Model Users” are those individuals that conform to the designers' intentions and that develop interpretations and behaviour that are consistent with the envisioned function of chatbots. However, users can have different or even contrary interpretations to the designers' intentions, thus finding the chatbot useless as unable to provide appropriate information they need to achieve their goals.

A cultural semiotic analysis can provide a better understanding of the real-life context of users, providing a photograph of the local cultural context in which chatbots operate, including contingent needs, local dialects and colloquialisms. The notion of Encyclopaedia by Eco is also useful to analyse the cultural context of users. It indicates the stock of shared signs that interpreters use during their interpretative processes. At the global level, the Encyclopaedia contained all the potential interpretations circulating in culture. At local levels, there was the routinised set of instructions to interpret specific portions of the socio-cultural space (Eco 1984: 68). He called this set of instructions “encyclopaedic competence” (Eco 1984: 2–3). Local cultures could select relevant portions of knowledge to delimit their own areas of consensus and thus to differentiate themselves from other cultures (Hajer & Wagenaar 2003: 27). Analysing the culture of users could mean to have insight into “a number of incorporated and (implicit) routinised ‘recurrent regularities’ about how to behave and act in specific situations”, including chatbot conversation (Othengrafen & Reimer 2013: 1273).

Drawing on this framework, an ethnographic approach based on semi-structured interviews or surveys to future users will be useful to analyse what people look for and how they ask for it during the conversation. Answering the following questions helps the building of the chatbot: 1) who are the target users of the chatbot? 2) What do they ask and what do they need? 3) How do they ask questions? Regarding the latter, a request can be expressed in almost unlimited ways and it is impossible to map every single utterance users may use. Still, specific groups of users share recurrent syntactic or morphological structures and sociolinguistic characteristics that have to be taken into account while implementing the linguistic features of the chatbot. Conversational AI platforms belong to the dominant IECR paradigm (section 2.2) and are therefore supported by NLU algorithms. This means that all the possible linguistic formulations of a request are not needed, because they can generalise and extract the underlying intent trained on a limited set of examples.

c. Create the chatbot’s personas: chatbot personas refer to the identity, tone and language style of chatbot. It aims to build stronger personal connections with the users, adding fun, clarity and smoothness to the conversation. As the identity of social groups is built on differences

and similarities, also the personas of chatbots is relational: it is built in relation to personality types circulating in the cultural context and opposed to competitors. Chatbot personas are not built once and for all: designers should continuously optimise conversational flows, as users' needs and their way to solve them constantly change. Semiotics, jointly with ethnosemiotic methods (Marsciani 2007), allows designers to better understand personas in relation to the cultural context, while focusing on meaning, narrative and the ways to engage with the audience culturally. Three aspects are important in tailoring chatbot personas:

- *Personality*: chatbot personas can be formal or informal, serious or fun, enthusiastic or dispassionate. Defining in advance chatbot personas helps to decide how the chatbot talks and acts in conversation. It influences its identity, tone of voice and the conversational flow. Chatbot should speak the same language of the target users. In relation to this, Bennett (2018) suggests to expressively manipulate orthography to define the degree of considerateness or enthusiasm the chatbot is supposed to express. During the conversation, humour could be used as a tool to improve the user experience (Jain *et al.* 2018). It allows users to establish an emotional connection with the conversational experience. Humorous responses should be calibrated in relation to the goals of the chatbots. Since efficiency and productivity are what users expect from a chatbot itself (Brandtzaeg & Følstad 2018), humorous responses can be counterproductive. Surprising effects could instead be used in purely small-talk conversations, in which the chatbots does not have to pursue the goal it has been mainly developed for, but entertain the users and build truthfulness. Another important issue is the attribution of a gender to a chatbot: when the chatbot is anthropomorphised, the question is whether it should reproduce the traditional conceptions of gender, including gender biases and issues, or it could be seen as an opportunity to deconstruct prejudices and deformed expectations. The UNESCO for the EQUALS Skills Coalition 2019 pictures a prevalent tendency to give female voices to voice assistants, thus reproducing subordinations and servant dynamics as well as the tolerance towards sexist verbal harassment (West *et al.* 2019).

- *Job description*: setting a job description for the chatbot may improve the building phase. Does the chatbot have a technical background or is a member of the customer support team? Is it an expert in marketing or in sales? Trying to write down the competences that are required for the chatbot's imaginary job position and the task it has to execute is a good way to stay consistent and identify jargon and technical language it has to manage (Moore & Arar 2019).
- *Self-awareness*: even though a chatbot of course has no real self-perception in a philosophical sense, it is still able to speak in some way. This is the reason why it should be able to formulate sentences about itself, managing topics like what a chatbot is and what are its main competencies, and it does not have. This awareness may help the user to navigate the conversational space created by the chatbot and give them the possibility to formulate questions and requests (Følstad & Brandtzaeg 2020).

4. Conclusions

This paper proposed a semiotic and linguistic approach to contribute to chatbot design. This approach can prove useful to focus on the three dimensions of chatbots: a) the user/chatbot goal; b) the target user, her context and interpretations; c) the chatbot personas. Specifically, a semiotic perspective can identify not only which information chatbots provide to users and how, but also how users make and interpret meanings in conversation, considering the particular cultural context in which they occur. In this paper, the scope was limited to present the theoretical and methodological foundation of the semiotic and linguistic approach to building chatbot. Future research will operationalise the framework, testing its components on real-life cases.

References

- ADAMOPOULOU E., MOUSSIADES L. (2020), An Overview of Chatbot Technologies. *Artificial Intelligence Applications and Innovations*, 584: 373–83.

- BALA K., KUMAR M., HULAWALE S., PANDITA S. (2017), Chat-Bot for College Management System Using A.I. *International Research Journal of Engineering and Technology*, 4 (11): 26–30.
- BENNETT G.A. (2018), Conversational style: Beyond the nuts and bolts of conversation. In R.J. MOORE R., ARAR H., SZYMANSKI M., & G. REN G. REN, eds, *Studies in Conversational UX Design*, 161–180. Springer, International Publishing.
- BRANDTZÆG P.B., FØLSTAD A. (2018), Chatbots: Changing User Needs and Motivations. *Interactions* 25 (5): 38–43.
- BRAUN D., MENDEZ A.H., MATTHES F., LANGEN M. (2017), Evaluating natural language understanding services for conversational question answering systems. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 174–185).
- BRAUN D., MATTHES F. (2019), Towards a Framework for Classifying Chatbots. In *Proceedings of the 21st International Conference on Enterprise Information Systems*, 1, 484–489.
- BROWN G., YULE G. (1983), *Discourse Analysis*. Cambridge: Cambridge University Press.
- BUBLITZ W., NORRICK N.R. (2011), Introduction: The Burgeoning Field of Pragmatics. In W. BUBLITZ & N.R. NORRICK, eds, *Foundations of Pragmatics*, 1–20. Berlin: Mouton de Gruyter.
- CERCAS CURRY A., HASTIE H., RIESER V. (2017), A Review of Evaluation Techniques for Social Dialogue Systems. In *Proceedings of 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents*, 25–26.
- CHAKRABARTI C., LUGER G.F. (2015), Artificial Conversations for Customer Service Chatter Bots: Architecture, Algorithms, and Evaluation Metrics. *Expert Systems with Applications*, 42 (20), 6878–6897.
- CHAVES A.P. GEROSA M. (2019), How Should My Chabot Interact? A Survey on Human–Chatbot Interaction Design. *arXiv:1904.02743v1 [cs.HC]*.
- CHAVES A.P., DOERRY E., EGBERT J., GEROSA M. (2019), It's How You Say It: Identifying Appropriate Register for Chatbot Language Design. In *Proceedings of the 7th International Conference on Human–Agent Interaction*, 102–9.
- CHEEPEN C. (1988), *The Predictability of Informal Conversation*. London: Pinter.
- CIECHANOWSKI L., PRZEGALINSKA A., MAGNUSKI M., GLOOR P. (2019), In the Shades of the Uncanny Valley: An Experimental Study of

- Human–Chatbot Interaction. *Future Generation Computer Systems*, 92, 539–548.
- CLARK H.H. (1996), *Using language*. Cambridge University Press.
- CLARK L., PANTIDI N., COONEY O., DOYLE P., GARAIALDE D., EDWARDS J., SPILLANE B., GILMARTIN E., MURAD C., MUNTEANU C., WADE V., COWAN B.R. (2019), What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. *arXiv:1901.06525v1 [cs.HC]*.
- COHEN M.H., COHEN M.H., GIANGOLA J.P., BALOGH J. (2004), *Voice user interface design*. Addison–Wesley Professional.
- COLBY K.M., WEBER S., DENNIS F.H. (1971), Artificial Paranoia. *Artificial Intelligence* 2 (1): 1–25.
- DALE R. (2016), Industry Watch. The Return of the Chatbots. *Natural Language Engineering*, 22 (5), 811–817.
- DASGUPTA R. (2018), *Voice User Interface Design. Moving from Gui to Mixed Modal Interaction*. New York: Apress.
- ECO U. (1979), *The Role of the Reader*. Bloomington: Indiana University Press.
- (1984), *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press.
- (1990), *The Limits of Interpretation*. Bloomington: Indiana University Press.
- FADHIL A., SCHIAVO G. (2019), Designing for Health Chatbots. *arXiv pre-print arXiv:1902.09022*.
- FRYER K.L., CONIAM D., CARPENTER R., LĂPUȘNEANU D. (2020), Bots for Language Learning Now: Current and Future Directions. *Language Learning & Technology*, 24(2), 8–22.
- FØLSTAD A., SKJUVE M., BRANDTZAEG P.B. (2019), Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. In S. BODRUNOVA O., KOLTSOVA A., FØLSTAD H., HALPIN P., KOLOZARIDI L., YULDASHEV A., SMOLIAROVA, & H. NIEDERMAYER, eds., *Internet Science. INSCI 2018. Lecture Notes in Computer Science*, vol 11551, 145–156. Springer, Cham.
- FØLSTAD A., BRANDTZAEG P.B. (2020), Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience*, 5(1), 1–14.
- GNEWUCH U., MAEDCHE A., MORANA S. (2017), Towards Designing Cooperative and Social Conversational Agents. In *Proceedings of the 38th International Conference on Information Systems*, 1–13.

- GREIMAS A.J. (1966), *Sémantique Structurale*. Paris: Larousse.
- (1970), *Du Sens*. Paris: Seuil.
- (1983), *Du Sens II*. Paris: Seuil.
- GRICE H.P. (1975), Logic and Conversation. In P. COLE & J.L. MORGAN, eds., *Syntax and Semantics*, Vol. 3, Speech Acts, 41–58. New York: Academic Press.
- HAJER M., WAGENAAR H. (2003), *Deliberative Policy Analysis: Understanding Governance in the Network Society*. Cambridge: Cambridge University Press.
- HILL J., FORD W.R., FARRERAS I.G. (2015), Real Conversations with Artificial Intelligence: A Comparison between Human–Human Online Conversations and Human–Chatbot Conversations. *Computers in human behavior* 49, 245–250.
- HOY M.B. (2018), Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly* 37 (1), 81–88.
- HUSSAIN S., AMERI SIANAKI O., ABABNEH N. (2019) A Survey on Conversational Agents/Chatbots Classification and Design Techniques. In: L. BAROLLI M., TAKIZAWA F., XHAFI & T. ENOKIDO, eds., *Web, Artificial Intelligence and Network Applications*. WAINA 2019. *Advances in Intelligent Systems and Computing*, vol. 927. Berlin: Springer.
- JACQUES R., FØLSTAD A., GERBER E., GRUDIN J., LUGER E., MONROY–HERNÁNDEZ A., WANG D. (2019), Conversational Agents: Acting on the Wave of Research and Development. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*.
- JACQUET B., BARATGIN J., JAMET. F. (2018), The Gricean Maxims of Quantity and of Relation in the Turing Test. In *Proceedings of the 11th International Conference on Human System Interaction — HSI 2018*, 332–338.
- JAIN M., KUMAR P., KOTA R., PATEL S.N. (2018), Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference*, 895–906.
- JEFFERSON G. (1978) Sequential Aspects of Storytelling in Conversation. In J. SCHENKEIN, ed., *Studies in the organization of conversational interaction*, 219–248. New York: Academic Press.
- JOVANOVIC M., BAEZ M., CASATI F. (2020), Chatbots as Conversational Healthcare Services, *arXiv:2011.03969*.
- JURAFSKY D., MARTIN H.J. (in press) (2021), *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed. draft).

- KERBRAT-ORECCHIONI C. (1997), A Multilevel Approach in the Study of Talk-in-Interaction. *Pragmatics*, 7 (1), 1–20.
- KLIMCZAK P., WIRSCHING G., GRABEN P. (2020), Machine Semiotics. *ArXiv abs/2008.10522*.
- KLIMCZAK P., WIRSCHING G. (2020), Machine Semiotics. *arXiv preprint arXiv:2008.10522*.
- LEVINSON S.C. (1983), *Pragmatics*. Cambridge: Cambridge University Press.
- LINELL P., LUCKMANN T. (1991), Asymmetries in Dialogue: Some Conceptual Preliminaries. In I. Markova & K. Foppa, *Asymmetries in dialogue*, 1–20. New York: Harvester Wheatsheaf.
- LUGER E., SELLEN A. (2016), “Like having a really bad PA”: The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.
- MARKET DATA FORECAST. (2020), Chatbot Market. August 2020. Available at: <https://www.marketdataforecast.com/market-reports/chatbot-market>, accessed 16.5.2021.
- MARSCIANI F. (2007), *Tracciati di Etnosemiotica*. Milan: Franco Angeli.
- MATHUR V., SING A. (2018), The Rapidly Changing Landscape of Conversational Agents. *arXiv:1803.08419v2 [cs.AI]*.
- MAULDIN M. (1994), ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, AAAI Press, 16–21.
- MCTEAR M., CALLEJAS Z., GRIOL D. (2016), *The Conversational Interface. Talking to Smart Devices*. Berlin: Springer.
- MCTEAR M. (2020), *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*, Williston: Morgan & Claypool.
- MOORE R.J., ARAR R. (2019), *Conversational Ux Design. A Practitioner’s Guide to the Natural Conversation Framework*. Association for Computing Machinery.
- OTHENGRAFEN F., REIMER M. (2013), The Embeddedness of Planning in Cultural Contexts Theoretical Foundations for the Analysis of Dynamic Planning Cultures. *Environment and Planning A*45. 1269–1284.
- PANESAR K. (2019), Conversational Artificial Intelligence (AI) — Demystifying Statistical vs Linguistic NLP Solutions. *Journal of Computer-Assisted Linguistic* 4, 47–79.

- PEARL C. (2016), *Designing Voice User Interfaces. Principles of Conversational Experiences*. Sebastopol: O'Reilly.
- RADZIWIŁ N.M., BENTON M.C. (2017), Evaluating Quality of Chatbots and Intelligent Conversational Agents. *arXiv 1704.04579*.
- REEVES B., NASS C. (1996), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People*. Cambridge University Press.
- SACKS H. (1984), Notes on Methodology. In J.M. Atkinson and J.C. Heritage, eds., *Structures of Social Action: Studies in Conversation Analysis*, 21–27. Cambridge: Cambridge University Press.
- SACKS H., SCHEGLOFF E.A., JEFFERSON G. (1974), A Simplest Systematics for Organization of Turn-Taking for Conversation. *Language* 50: 696–735.
- SACKS H., SCHEGLOFF E.A. (2007), Two Preferences in the Organization of Reference to Persons in Conversation and their Interaction. In N.J. ENFIELD & T. STIVERS, eds., *Person Reference in Interaction: Linguistic, Cultural and Social Perspectives, Language Culture and Cognition*, 23–28. Cambridge University Press.
- SHAWAR B.A., ATWELL E. (2007), Chatbots: Are they Really Useful?. *LDV-Forum* 22(1), 29–49.
- SCHEGLOFF E.A. (1968), Sequencing in Conversational Openings. *American Anthropologist* 70, 1075–1095.
- (1999), Discourse, Pragmatics, Conversation, Analysis. *Discourse Studies* 1, 405–35.
- (2007), *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Vol. I. Cambridge: Cambridge University Press.
- SCHEGLOFF E.A., SACKS H. (1973), Opening up Closings. *Semiotica* 8: 289–327.
- SCHEGLOFF E.A., JEFFERSON G., SACKS H. (1977), The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53 (2), 361–382.
- SEARLE J.R. (1969), *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- TSVETKOVA M., GARCÍA-GAVILANES R., FLORIDI L., YASSERI T. (2017), Even Good Bots Fight: The case of Wikipedia. *PLoS one*, 12(2).
- TURING A. (1950), Computing Machinery and Intelligence. *Mind*, LIX: 433–460.

- VALÉRIO F.A., GUIMARÃES T.G., PRATES R.O., CANDELLO H. (2017), Here's What I Can Do: Chatbots' Strategies to Convey their Features to Users. In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, 28.
- VALTOLINA S., BARRICELLI B.R., DI GAETANO S. (2020), Communicability of Traditional Interfaces Vs Chatbots in Healthcare and Smart Home Domains. *Behaviour & Information Technology* 39 (1), 108–132.
- WALLACE R.S. (2003), *The Elements of Aimpl Style*. A.L.I.C.E. Artificial Intelligence Foundation Inc.
- WEST M., KRAUT R., EI CHEW H. (2019), *I'd blush if I could: closing gender divides in digital skills through education*.
- WINKLER R., SOELLNER M. (2018), Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. In *Academy of Management Proceedings*, 1.
- WU W., GUO Z., XIANGYANG Z., HUA H., XIYUAN. Z., RONGZHONG L., HAIFENG W. (2019), Proactive Human-Machine Conversation with Explicit Conversation Goals. *arXiv:1906.05572v2 [cs.CL]*.
- WU, Y., WU W., XING C., ZHOU M., LI Z. (2017), Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 496–505, Vancouver, Canada. Association for Computational Linguistics.
- WUENDERLICH N.V., PALUCH S. (2017), A Nice and Friendly Chat with a Bot: User Perceptions of AI-Based Service Agents. *ICIS 2017 Proceedings*, 1–12.
- WEIZENBAUM J. (1966), ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM*, 9(1), 36–45.
- ZUE V.W., GLASS J.R. (2000), Conversational Interfaces: Advances and Challenges. In *Proceedings of the IEEE*, 88 (8), 1166–1180.

L'INTELLIGENZA DELLE I.A. COME EFFETTO DI SENSO SEMIO-NARRATIVO NEI GIOCHI DIGITALI UNA RIVOLUZIONE SEMIOSOFICA

GIANMARCO THIERRY GIULIANA*

ENGLISH TITLE: *A.I. intelligence as a semio-narrative effect in digital games. A semiosophical revolution.*

ABSTRACT: the question at the heart of this contribution concerns the impossibility for a human agent to distinguish itself from an AI on the basis of its performative behavior exhibited in a video game and the socio-cultural implications that this impossibility entails. Thus, this is a work that investigates the effect in light of which we ascribe to the AIs present in most common video games the quality of being intelligent. This investigation begins with a semantic observation on the performative implications that are present in the very definition of the notion of “intelligence” in many dictionaries. Implication we point to as an ideological connotation insofar as the intelligence characteristic of the human species seems to be able to be reduced to the narrative demonstration of knowing how to do and doing well. We then turn from dictionaries to social discourses about digital games and observe under what conditions A.I.’s create a sense effect of stupidity and under what conditions they appear to be intelligent. We then focus on the latter and semiotically explain this attribution on the basis of two different ways in which it manifests itself in texts: a narrative-mimetic construction and a pragmatic one. In the first case we describe a mirror situation between human agent and machinic agent, such that in fact the agency of the A.I. is naturally interpreted in narrative terms through the categories employed by the human agent itself. In the second case, we analyze the different technologies behind the demonstrated effectiveness of A.I. (F.S.M, G.O.A.P, B.N.N., Utility A.I., Reinforcement Learning, Behavior Tree) in some iconic video games (Pac-Man, Half-Life, Fear, Alien Isolation, Black & White, Forza) and show how these systems are built on a modelization of performative human action viewed from different aspects (effective problem solving, adaptivity, multiplicity of options, contextualization, potentiality, inferential construction of a model opponent, and experience). Thus demonstrating an almost impossible distinction between A.I. and human within a performative framework, we argue in our conclusions that digital games are an important part of contemporary social discourse around the unicum of human intelligence and propose two alternatives to a functional-performative conception of it: an environmental and a ludosemiotic conception.

KEYWORDS: semiotics, artificial intelligence, intelligence, digital games, narration.

*Università degli Studi di Torino.

1. Introduzione

Nel 1950 il matematico Alan Turing concepisce un “gioco imitativo” (Turing 1950) il cui scopo è quello di permettere ad una macchina di apparire potenzialmente come un essere umano sulla base del suo comportamento. Da allora, tanto il tema dell’inganno macchinico quanto quello della potenziale umanità della macchina sono stati al centro di innumerevoli discorsi e narrazioni. Questa costellazione di testi nella semiosfera ha così tratto da diverse prospettive la questione del proprio dell’intelligenza umana e la ha continuamente riarticolata in funzione del costante progresso tecnico delle I.A. Un processo che, come abbiamo dimostrato in un precedente lavoro (Giuliana 2021), ha portato a una estensione semantica della nozione di *persona* nella semiosfera contemporanea.

Esiste un contesto, tuttavia, in cui questo confronto fra l’uomo e la macchina ha assunto un significato molto particolare: i giochi digitali. Il motivo principale di ciò risiede nel fatto che, a differenza dei numerosi ambiti della vita sociale in cui siamo abituati a delle intelligenze artificiali create per servirci, nei videogiochi le intelligenze artificiali sono nate con uno scopo principalmente *oppositivo*. Infatti, le I.A dei giochi hanno definito il medium videoludico proprio in quanto, sin dalla sua nascita commerciale, esse avevano lo scopo di sostituire un ipotetico “giocatore 2” avversario del primo. Certo, da un punto di vista generale questa opposizione è pur sempre funzionale al divertimento del giocatore. Ciò non toglie però che sin dagli anni ‘70 i videogiocatori umani si sono visti opposti alle I.A., si sono confrontati con esse e sono stati spesso sconfitti da queste intelligenze artificiali contro la loro volontà.

Alla luce di questa specificità, i mondi virtuali dei giochi digitali sono un contesto molto interessante per studiare il senso sociale dell’intelligenza artificiale e la rivoluzione *semiosofica* che essa ha comportato. Quale rivoluzione? L’abbandono di una retorica performativa dell’intelligenza *tout court* secondo cui essere intelligenti vuol dire innanzitutto *saper e poter fare*. Al fine di difendere questa idea, in questo contributo vogliamo in qualche modo capovolgere il quesito di Turing chiedendoci se è possibile per un giocatore umano dimostrare di *non essere* una intelligenza artificiale in una data realtà ludico-virtuale. La

dimostrazione di questa impossibilità avviene in questo contributo in quattro diverse sezioni mirate a fare un confronto fra i processi e le funzioni decisionali e agentivi delle I.A. e quelli degli esseri umani dal punto di vista di come essi vengono interpretati e di conseguenza mostrando come l'attribuzione dell'intelligenza a una agentività macchinica sia innanzitutto un effetto di senso semiotico.

Nella prima sezione, ciò avviene con una osservazione semantica sulle implicazioni performative che sono presenti nella definizione stessa della nozione di "intelligenza" in molti dizionari. Implicazione che indichiamo come connotazione ideologica nella misura in cui l'intelligenza caratteristica della specie umana sembra poter essere ridotta alla dimostrazione narrativa di un *saper fare* e un *fare bene*. Nella seconda sezione, mettiamo in luce come nei videogiochi di uso comune le I.A. possano sembrare tanto "stupide" quanto "intelligenti" ed esaminiamo in quali condizioni si verifica un giudizio o l'altro. Coerentemente con i nostri intenti, nella terza sezione ci interroghiamo su tutti quei casi in cui l'intelligenza dell'I.A. sembra assomigliare a quella umana. Specificamente, spieghiamo questo effetto di senso attraverso due diversi modi in cui esso viene costruito dai testi: una costruzione di tipo narrativo-mimetica e una di tipo pragmatica. Nel primo caso descriviamo una situazione a specchio fra agente umano e agente macchinico, tale per cui di fatto l'agency dell'I.A. viene naturalmente interpretata in termini narrativi attraverso le categorie impiegate dall'agente umano stesso. Nel secondo caso analizziamo le diverse tecnologie dietro l'efficacia dimostrata dalle I.A. in alcuni videogiochi iconici e mostriamo come questi sistemi siano costruiti su una modellizzazione di un agire umano performativo visto sotto sette aspetti diversi (problem solving efficace, adattività, molteplicità delle opzioni di adattamento, contestualizzazione, contestualizzazione efficace, costruzione inferenziale di un avversario modello ed esperienza). Dimostrando così una distinzione impossibile fra I.A. e umano entro un quadro performativo, nella nostra ultima sezione sosteniamo la tesi secondo cui i giochi digitali sono una parte importante del discorso sociale contemporaneo attorno all'unicum dell'intelligenza umana e proponiamo due alternative a una sua concezione funzionale-performativa: una concezione ambientale e una ludosemiotica.

2. Definizione performativa dell'intelligenza

Prima di ogni cosa, è necessario chiarire cosa intendiamo con “retorica performativa dell'intelligenza”. Con questa espressione, intendiamo la corrispondenza connaturata e connotata (Barthes 1964) nel linguaggio fra “essere intelligenti” e “essere competenti”. Una corrispondenza che troviamo nei dizionari di molte lingue:

Complesso di *facoltà* psichiche e mentali che *consentono* all'uomo di pensare, *comprendere* o *spiegare* i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e lo rendono insieme *capace di adattarsi* a situazioni nuove e di *modificare la situazione* stessa quando questa presenta ostacoli all'adattamento;” [...] Per estens., il *modo* intelligente, la perizia, l'ingegnosità con cui *si svolge una mansione* o *si compie un lavoro*: un articolo scritto con i., con molta i.; l'impianto è fatto con i.; condurre con i. le trattative; ha sempre eseguito con i. tutti gli incarichi affidatigli (treccani.it, corsivo nostro).

The ability to learn, understand and think in a logical way about things; *The ability to do this well* (oxfordlearnersdictionaries.com, corsivo nostro).

Aptitude d'un être humain à *s'adapter* à une situation, à *choisir des moyens d'action* en fonction des circonstances [...]. Synonymes: adresse – doigté – *maestria* – *maîtrise* — *virtuosité* (larousse.fr, corsivo nostro).

In tutti questi esempi vediamo come l'intelligenza sia definita in quanto capacità *tipicamente umana* di comprendere la realtà e di applicare dinamicamente questa comprensione in modo funzionale alle sue intenzioni e azioni. Una definizione basata su una ideologia antropocentrica che presuppone l'uomo come superiore tanto nel pensare quanto nell'agire (e talvolta persino nel sentire) e da cui deriva ciò che chiamiamo “retorica performativa dell'intelligenza”.

Da un punto di vista semiotico, questa retorica può essere concepita sia attraverso un modello testualista che attraverso un modello interpretativo. Secondo il primo modello, l'intelligenza può essere dimostrata come la capacità per un *Soggetto* di ottenere un *Oggetto di Valore* conteso con un *Anti-Soggetto* attraverso una *Performance* che

presupponga una doppia *modalizzazione* del *Saper Fare* e del *Poter Fare* al fine di ottenere una *Sanzione* positiva. Da un punto di vista interpretativo, invece, questa retorica dell'intelligenza sarà dimostrata come la capacità per un *Interprete* di formulare *Inferenze* corrette rispetto a una data *Scommessa* interpretativa basandosi su *Scenari/Script* preesistenti e su una *Enciclopedia* di conoscenze fondamentali per il costruirsi di un *Avversario Modello* (Eco 1979, p.54) verosimile.

In entrambi i casi, potremo parlare di una “retorica performativa dell'intelligenza funzionale ad una ideologia antropocentrica” tutte le volte che all'interno di un certo discorso viene postulata *a priori* una maggiore probabilità di esito positivo se il Soggetto/Interprete possiede la marca dell'umano ed è opposto ad una agency che è invece di natura non-umana.

3. Narrazione e gioco dell'I.A.

La specificità delle narrazioni incentrate sul tema dell'intelligenza artificiale è proprio di mettere in crisi questa retorica. Questo sia rappresentando robot capaci di fare cose tipicamente associate agli esseri umani che mettendo in discussione la qualità stessa del pensare e performare umano, pensiamo per esempio al dialogo nel fim *I, Robot* (Proyas 2004) fra l'agente umano Del Spooner e l'androide Sonny:

«Tu sei solo una macchina, una imitazione della vita. Un robot può scrivere una sinfonia? Un robot può trasformare una tela bianca in un'opera, un capolavoro?»

«Lei può farlo?»

Se nelle narrazioni ciò avviene però attraverso appunto delle storie fittizie che concepiscono molto spesso delle IA irrealistiche nella loro complessità (e cioè di fatto attraverso un'altra forma retorica), questa stessa retorica performativa dell'intelligenza viene messa in crisi in modo decisamente più pragmatico all'interno del mondo ludico in cui le intelligenze artificiali con cui si confronta il giocatore sono fattuali. Momenti come la sconfitta di Kasparov contro Deep Blue nel 1997 al

gioco degli scacchi o come la sconfitta di Lee–Sedol contro AlphaGo nel 2016 hanno segnato non solo la storia dei giochi e dell’informatica, ma anche l’immaginario sociale su ciò che è davvero lo specifico dell’essere umano e della sua intelligenza.

Sia Deep Blue che AlphaGo sono tuttavia delle intelligenze artificiali avanzatissime create appositamente per sconfiggere una specie di “ideale performativo” dell’intelligenza umana sul terreno di giochi estremamente complessi. È possibile dire però lo stesso del confronto fra uomo e macchina all’interno dei videogiochi? Per rispondere a questa domanda bisogna prima di tutto decidere di quale tipo di I.A. interessarsi. Le intelligenze artificiali odierne, infatti, possono avere diversi compiti fra cui quello di creare il mondo del giocatore (con la tecnica nota come *PCG*), di costruire situazioni narrativamente interessanti o ancora di gestire il traffico delle automobili. In questo articolo noi ci concentreremo unicamente su quelle I.A. dei giochi di massa che sono in aperta *competizione* col giocatore o che *collaborano* esplicitamente con lui.

Queste intelligenze artificiali presenti nei videogiochi pensati per un “pubblico di massa”, infatti, sono decisamente più semplici (non create in laboratori avanzatissimi e con milioni di fondi alle spalle) e agiscono all’interno di giochi apparentemente più semplici. Entro le mura di questa semplicità, un primo sguardo alle comunità di videogiocatori sembrerebbe piuttosto confermare l’idea che le realtà virtual–ludiche siano un luogo in cui gli esseri umani possono confortevolmente consolidare il loro *bias* antropocentrico di superiorità. I primi esempi che vengono qui in mente sono i personaggi non giocabili (NPCs) come i mercanti, vere e proprie marionette che in un gioco come *Skyrim* (Bethesda 2011) controlleranno visivamente che il giocatore non rubi nulla dal negozio ma si lasceranno serenamente mettere un secchio in testa cosicché, ormai cieche, il giocatore possa derubare ogni cosa in tutta tranquillità. Questa tipologia di agenti videoludici è tuttavia un caso a parte di cui, di fatto, è molto difficile parlare di intelligenza o stupidità, poiché essi sono chiaramente fuori dalle logiche di vittoria e sconfitta che caratterizzano la maggior parte dei videogiochi. Vi sono però altre tipologie di agenti che rientrano perfettamente nella categoria delle “I.A. stupide” e di cui alcuni esempi sono addirittura iconici della cultura dei *gamers*.

Natalya in *007 Golden Eye* (Rare 1997), i compagni della serie *Fallout*, gli amici in *Tales of Destiny* (Wolf Team 1997), il partner in *The Uncanny X-Men* (LJN 1989), Sheva in *Resident Evil 5* (Capcom 2009), i marines alleati nella saga di *Halo*, i superstiti in *Dead Rising* (Capcom 2006) sono solo alcuni fra i più noti esempi. Persino interi giochi come *Daikatana* (Ion Storm 2000) sono stati compromessi da pessime I.A.

Tutti gli esempi iconici sopracitati hanno però una caratteristica semanticamente importante, ovvero sono I.A. che sono *alleate* del giocatore e la cui intelligenza è dunque valutata dai giocatori sulla base di un riflesso del loro agire e ragionare (e cioè dei loro PN e delle loro inferenze) in *una stessa situazione*. Ovviamente vi sono anche casi in cui l'indiscutibilità performativa dell'intelligenza umana viene rappresentata in situazioni oppostive, e così per esempio nello sparatutto multigiocatore *CS:GO* (Valve 2012) farsi uccidere da un bot può essere motivo di derisione. O ancora, nel 2020 il battle royale *PUBG* (Bluehole 2017), uno fra i videogiochi più giocati al mondo, ha visto una enorme protesta da parte della comunità dopo l'introduzione di bot avversari terribilmente "stupidi", al punto che sono nate delle vere e proprie parodie. Questi casi sono tuttavia meno comuni e socialmente significativi per due diversi motivi. Il primo è che in alcuni generi particolarmente "ludici", come il musou (la serie *Dinasty Warriors* di Koei) o il sandbox (la serie *GTA* di Rockstar), la "passività" e "semplicità" delle agency delle intelligenze artificiali è un valore *necessario* al divertimento del giocatore dandogli un senso di "empowerment". Il secondo è che la posizione oppostiva di questi agenti è spesso *asimmetrica* rispetto al giocatore. Così un umano difficilmente farà un paragone fra la sua intelligenza e quella di un avversario comune (*mob*) che rappresenta una sfida non in quanto singolo ma in quanto parte di un gruppo o di una certa situazione. L'effetto di intelligenza dell'I.A., in termini greimassiani, non riposa dunque sul singolo tipo e sulla singola occorrenza di attore-oppositore (esempio: il funghetto di *Super Mario Bros*) ma sull'insieme di strategie e capacità che questi *dimostrano aspettualmente* in quanto attori collettivi (Greimas e Landowski 1976) della "mente" dell'I.A. che è invece il vero e proprio Anti-Soggetto del testo.

Ecco perché gli agenti oppositivi vengono spesso indicati dalla comunità dei giocatori come manifestazioni del *meglio* dell'intelligenza

artificiale nei videogiochi, con alcuni esempi iconici che esamineremo nella prossima sezione. Ancora più interessante, infine, è per noi l'osservazione di come questa ricorrente attribuzione di intelligenza ad alcune delle I.A. oppostive non possa essere ricondotta ad una effettiva differenza di programmazione. Infatti, è noto come durante i playtest del gioco *Halo* (Bungie 2001) ai giocatori sia stato chiesto quali fra due nemici guidati dall'I.A. fosse più intelligente, e come i giocatori abbiano espresso una chiara preferenza nonostante in realtà le I.A. fossero identiche e solamente i punti di vita fossero differenti.

Ciò che è interessante da un punto di vista semiotico, in definitiva, è dunque come l'intelligenza umana riconosciuta nell'I.A. sia di fatto *un effetto di senso* determinato tanto da fattori legati alla narratività quanto all'interpretazione dietro una agentività, e ciò proprio in virtù dell'identificazione dell'intelligenza umana con una performance concepita in termini narrativi. Alla luce di ciò, nella prossima sezione cercheremo dunque di esaminare i diversi modi in cui questa somiglianza semio-narrativa si costruisce nelle realtà ludico-virtuali.

4. Effetto Mimetico ed Effetto Pragmatico dell'intelligenza delle I.A.

In che modo le I.A. presenti nei videogiochi sembrano intelligenti agli occhi degli esseri umani, a volte persino al punto da far svanire la specificità dell'agire umano? Nelle sezioni precedenti abbiamo messo in evidenza la natura intrinsecamente semiotica di una qualsiasi dimostrazione performativa sulla base della quale è possibile attribuire a un dato agente la qualità di essere intelligente. Partendo da questo, diremo che *l'effetto di senso di intelligenza delle I.A. nella realtà ludico-virtuali si dà sulla base di una modellizzazione simil-uguale del comportamento esibito dalla macchina*. In altre parole, questo effetto di senso nasce quando l'osservatore umano interpreta l'agire macchinico attraverso lo stesso modello che egli usa per mettere in atto il proprio pensiero-azione.

Questa modellizzazione ha due diverse e complementari facce. La prima è di tipo narrativo-mimetico e consiste nel creare una situazione a specchio fra l'agire dell'umano e l'agire dell'I.A.: banalmente nel dare a entrambi lo stesso spettro di azioni possibili e nel rappresentarle

contemporaneamente sullo schermo. In uno sparatutto, per esempio, le macro-opzioni a disposizione del giocatore per superare le difficoltà si possono ridurre a qualcosa come “cercare” “attaccare” “mettersi al riparo” o “fuggire”, azioni-tipo che (come vedremo in 3.3) sono le stesse categorie utilizzate negli alberi decisionali delle I.A. Dato il *surplus di senso* che caratterizza il pensare umano (Volli 2005), tutte le volte che queste I.A. esibiscono un comportamento simile a quello del giocatore è assolutamente naturale per lui interpretare queste azioni secondo delle logiche narrative che gli sono proprie. Così, per esempio, tutte le volte che una I.A. sembra agire al fine di preservare sé stessa dai colpi dell'umano, essa sembra narrativamente condividere lo stesso valore della sopravvivenza e della vittoria che caratterizza il giocatore. Quando poi in un gioco come *Bioshock* (Bioware 2007) un NPC umanoide non soltanto si cura dai colpi ricevuti ma lo fa attraverso lo stesso oggetto che usa il giocatore, l'effetto mimetico raggiunge il suo apice suggerendo l'idea che I.A. e giocatore condividano persino le stesse regole. Inoltre, nel momento in cui le I.A. esibiscono queste azioni in una *sequenza* (es: sparare, scappare, mettersi al riparo, sparare nuovamente) è possibile per il giocatore modellizzare narrativamente l'agire macchinico tanto in termini di *sceneggiature* (Eco 1979) che in termini di *Programmi Narrativi* primari e secondari (Greimas 1970). Ciò spiega semioticamente il fatto che in *Halo* agenti con una stessa e identica I.A. ma con diverse capacità di sopravvivenza vengano valutati come diversamente intelligenti dagli osservatori umani: banalmente quelli con meno punti vita hanno meno possibilità di eseguire sequenze di azioni narrativamente sensate e interpretate sotto la forma di un “piano”.

La seconda faccia della modalizzazione, complementare con la prima, è invece di natura pragmatica in quanto permette all'umano di spiegare l'*efficacia* dell'I.A. sulla base di capacità (*competenze modali*) che sono proprie del giocatore in carne ed ossa. In un qualsiasi gioco, infatti, ogni decisione e azione intrapresa al fine di superare una data difficoltà rappresenta un problema al tempo stesso logico (capire il problema e individuare una soluzione fra molte possibili), inferenziale (concepire i possibili esiti della soluzione scelta e i criteri delle eventuali contromosse future dell'agency avversaria), percettivo (usare le informazioni sensoriali tanto per comprendere il problema che per risolverlo), motorio (eseguire una

data successione di input) e cognitivo (compiere tutto entro il minor lasso di tempo possibile e usando la propria memoria). La dimostrazione performativa dell'intelligenza nel contesto di un videogioco viene dunque misurata rispetto a queste capacità *implicate nell'agency efficace*. Non si tratta più qui dunque solamente di una interpretazione dell'agire dell'I.A. a foce secondo un modello narrativo umano-tipico, ma piuttosto di spiegare la modellizzazione narrativo-mimetica riconoscendo fattualmente e funzionalmente nell'I.A. una modellizzazione a monte (e cioè nella programmazione) del modo umano di pensare e agire *bene*.

Nel contesto del videoludico, questo pensare e agire intelligente si può manifestare in sette modi:

1. L'agire efficace
2. L'agire adattivo
3. L'agire adattivo molteplice
4. L'agire adattivo contestuale
5. L'agire adattivo contestuale-potenziale
6. L'agire inferenziale
7. L'agire esperienziale

Nelle sottosezioni seguenti esamineremo nel dettaglio ognuno di essi e mostreremo in che modo le tecnologie informatiche in atto nei videogiochi siano di fatto capaci di emulare tutte queste forme di agire che, secondo una retorica antropocentrico-performativa dell'intelligenza, dovrebbero invece essere proprie solo dell'umano.

4.1. *L'agire efficace*

Vincere una partita in un videogioco vuol dire innanzitutto essere capaci di eseguire con successo le singole diverse azioni fallibili al suo interno. In un'ottica performativa dell'intelligenza, compiere *efficacemente* queste azioni "semplici" vuol dire "fare" prova di intelligenza. Prendiamo come esempio il noto gioco *Pac-Man* (Namco 1980) che richiede al giocatore di muoversi su una mappa "mangiando" tutti i vari pallini sullo schermo nel mentre evita di essere raggiunto da quattro fantasmini guidati da una I.A.. In un gioco simile, *sapere* valutare le distanze al fine di determinare il

miglior percorso possibile nello spazio è una azione semplice fondamentale. Ora, ognuno dei quattro fantasmmini si muove su quella mappa applicando continuamente una data formula matematica di calcolo del percorso “migliore” al fine di poter raggiungere il giocatore. Calcolo che è di certo la *soluzione migliore* ma di cui l'essere umano è di fatto incapace in quel lasso di tempo. Così, si arriva alla paradossale conclusione che se calcolare le distanze è un modo logico di pensare, e se calcolarle bene in un breve lasso di tempo vuol dire far prova di intelligenza, allora i fantasmi sono intelligenti, allora i fantasmi sono più intelligenti dell'uomo.

La conclusione paradossale appena esposta mette in evidenza come, essendo i videogiochi delle realtà innanzitutto matematiche ed avendo le I.A. un accesso direttamente matematico a quel mondo, non è possibile fare paragoni sensati fra intelligenza umana e macchinica sulla base delle azioni semplici. In un genere come lo sparatutto, per esempio, sia l'uomo che l'I.A. devono mirare e dare un input prima di sparare. Tuttavia, laddove l'uomo deve usare il suo sistema visivo e motorio per compiere queste azioni semplici, l'I.A. agisce in modo completamente diverso. Al punto che in un gioco come *Perfect Dark* (2000) un messaggio avverte il giocatore che alla massima difficoltà solo un errore informatico può far sì che un nemico manchi il suo colpo. Non di meno, proprio in virtù dell'esistenza culturale di una retorica performativa dell'intelligenza, queste azioni semplici sono una parte innegabile dell'effetto di intelligenza delle I.A. nei videogiochi: ecco perché i “Dark Sim” di *Perfect Dark* vengono descritti come esseri artificiali con una “*intelligenza quasi onnipotente*”.

4.2. *L'adattamento semplice*

Al fine di poter dimostrare il proprio dell'intelligenza umana secondo una logica performativa serve dunque un altro criterio, e un ottimo candidato sembra essere la capacità di adattamento che è comparsa diverse volte nei nostri dizionari. Esistono dunque I.A. abbastanza intelligenti da adattarsi alle azioni del giocatore umano così come egli si adatta all'agire delle I.A.? La risposta a questa domanda dipende principalmente da ciò che si intende con “adattamento”. Nella sua forma più semplice, l'adattamento è una *reazione* e da questo punto di vista già nel 1978 *Space Invaders* (Taito) presentava nemici che *rispondevano* al fuoco del

giocatore umano e *diventavano* sempre più aggressivi man mano che la vittoria del giocatore diveniva più probabile. Allo stesso modo, nel momento in cui un giocatore mangia un frutto in *Pac-Man*, i fantasmini passano da un comportamento di inseguimento a uno di fuga, simulando narrativamente un adattamento alla situazione e un loro attaccamento alla vita! Caratteristiche che, nuovamente, non permettono di distinguere *formalmente* l'agency umana da quella dell'I.A sulla base di un adattamento semplice e spiegano l'effetto di senso di intelligenza.

4.3. *L'agire adattivo molteplice*

Pur mostrando un qualche grado di adattamento, le navicelle di *Space Invaders* e i fantasmini di *Pac-Man* hanno un comportamento prestabilito estremamente semplice rispetto a quello del giocatore umano. Un criterio per una prova performativa dell'intelligenza umana diventa dunque il numero di opzioni a disposizione di una agency adattiva. È qui che entra in gioco la tecnologia algoritmica nota come "Finite State Machine" (FSM da ora in poi).

Le FSM hanno il compito di gestire il comportamento dell'I.A. in modo tale da far passare un agente da un certo *stato* a un altro in funzione di un dato *evento* attraverso ciò che viene chiamato una *transizione*. Ritornando all'esempio di 3.2, "inseguimento" e "fuga" sono due stati dei fantasmini che dipendono dagli evento del "frutto attivo" e "effetto frutto svanito". Col progredire della tecnologia le FSM sono tuttavia divenute più complesse e hanno così emulato ulteriormente l'agire molteplice che caratterizza il comportamento guidato dall'intelligenza umana. Fra i molti titoli possibili, *Half-Life* (1998) è certamente un esempio iconico. In questo gioco, ogni agente del mondo agisce sulla base di ottanta diversi stati-compiti (task) che sono azioni specifiche (es: "stare fermi", "inseguire" o "essere di fronte a") organizzate in *programmi* (*schedules*, per esempio "tendere una imboscata") sulla base di 32 diverse *condizioni* che sono di fatto delle *informazioni* sul mondo di ordine per lo più percettivo e che vengono costantemente aggiornate (es: nemico visibile, sparo sentito). L'agire dell'I.A. si fa così adattivo sulla base di una catena *predeterminata* di stati-azioni che organizza *dinamicamente* (in tempo reale) e realizza 5 "obbiettivi" (goal) generali

comuni a ogni agente (attaccare, muoversi, muoversi verso il nemico, mettersi al riparo, e nutrirsi) in una molteplicità di modi diversi e che aggiornano costantemente l'I.A. su quali programmi e compiti svolgere. Così il comportamento dell'I.A. presenta nel *suo comportamento apparente* delle variazioni che sono narrativamente coerenti con il pensare e l'agire dell'umano in quanto basato su una emulazione delle capacità percettive (vedere, sentire e persino odorare!) e logiche (se x allora y).

Facciamo un esempio per rendere tutto più chiaro: in un determinato luogo del mondo virtuale un giocatore umano spara verso un nemico guidato da una I.A., a quel punto come reagisce l'agente non-umano? Le possibilità sono molteplici. Un nemico non colpito sentirà lo sparo e userà questa informazione per 'localizzare' l'umano e 'dirigersi' verso di lui, diversamente un nemico 'ferito' da un giocatore che 'non percepisce' interromperà la sua 'routine' e tenderà a 'mettersi al riparo' e ad adoperarsi per 'individuarlo', mentre un nemico ferito da un giocatore che 'vede' tenderà a 'schivare' i suoi colpi o ad 'attaccarlo' e nel secondo caso a seconda della 'distanza' del giocatore opterà per un 'attacco ravvicinato' piuttosto che uno 'a distanza'.

Tutto questo non solo costituisce un agire molteplice in funzione delle capacità e informazioni dell'I.A., ma è funzionalmente simile all'agire del giocatore nello specifico contesto del videogioco.

4.4. *L'agire adattivo comune-potenziale*

Una volta escluso l'agire molteplice dalle manifestazioni proprie dell'intelligenza umana, si può a questo punto andare al principio decisionale dietro lo stesso agire per distinguere l'uomo dalla macchina. Infatti, le FSM, per quanto complesse, non pongono per l'I.A. nessun vero momento di dubbio e sono caratterizzate da una certa ripetitività proprio in virtù della loro natura sequenziale e predeterminata. Un essere umano, al contrario, non agisce naturalmente secondo sequenze di azioni "pronte all'uso" ma secondo un principio narrativo ed esaminando *contestualmente* tutte le possibilità in relazione al suo scopo. Sebbene le FSM siano basate su "eventi transitori" capaci di emulare in parte una certa agency contestuale, esse richiedono un enorme lavoro di scrittura sequenziale di ogni singolo agente del mondo a seconda del suo contesto predeterminato, il che è alla base del

loro principale difetto, ovvero la scalarità. Viceversa, la diversità decisionale dell'agire umano è significativa proprio perché si fonda su un modello cognitivo e comportamentale comune: la sua *modularità*. Sebbene per rimediare a questo limite le FSM siano state in parte migliorate (in *HFSM*), questo modo di pensare e agire è stato più specificatamente emulato dalle I.A. in due diversi modi: una variazione delle FSM chiamata "G.O.A.P" e un sistema alternativo noto come *Behavior Tree* (traducibile come "albero comportamentale", B.T. da ora in poi).

Nel primo caso, il gioco di riferimento è lo sparattutto *F.E.A.R* (Monolith 2005) in cui è stato implementato il sistema denominato *Goal-Oriented-Action-Planning*. Qui, gli obiettivi (goal) sono il punto di partenza *comune* del comportamento dell'I.A. e vengono selezionati contestualmente e dinamicamente secondo un *criterio di priorità*. Una volta selezionato un preciso obiettivo, a quel punto l'I.A. elaborerà un *piano* per svolgere al meglio quel compito sulla base di 120 possibili azioni, piano che sarà poi tradotto come una classica successione di stati. Infine, l'I.A. verificherà costantemente lo stato del piano e lo cambierà in caso di necessità.

Facciamo un esempio: l'I.A. ha selezionato l'obiettivo 'sparare al giocatore' come priorità in virtù del suo trovarsi nella stessa stanza e di tenere in mano un'arma da fuoco. Nel calcolare il modo migliore per realizzare questo *obbiettivo*, appare subito come *problema* da risolvere il fatto che il giocatore sia 'al riparo dietro un oggetto'. A quel punto l'I.A. determina, fra molteplici possibilità, che il *piano* migliore è quello di 'fiancheggiare' il giocatore e comincia ad eseguire una sequenza di azioni per farlo ('muoversi verso', etc.). Tuttavia all'improvviso il giocatore abbandona il suo riparo e corre ad armi spianate verso i nemici. Ciò ovviamente invalida il piano iniziale dell'I.A. e la spinge a creare, nel minor tempo possibile, un nuovo piano per cui per esempio saranno ora i soldati avversari a 'cercare un riparo' nelle vicinanze al fine di sopravvivere all'attacco del giocatore. La differenza semiotica principale rispetto al punto precedente sta nella natura *potenziale anziché predeterminata* di queste decisioni adattive, con comportamenti complessi (i piani) che emergono da azioni semplici e con una maggiore difficoltà del giocatore a prevedere il comportamento non-umano.

Una seconda alternativa alle FSM classiche sono i cosiddetti *behaviour trees* e sono una tecnologia molto diffusa nel contesto tecnologico dei

giochi digitali. I B.T. si presentano come veri e propri alberi decisionali *comuni a tutti gli agenti* che sono caratterizzati dalla definizione di *priorità* basate sul *contesto*. Sia il contesto che le priorità qui hanno lo scopo di *limitare le ramificazioni* dell'albero delle azioni possibili, rendendo l'I.A. più efficace nel breve termine e consentendole di esibire azioni più pertinenti dal punto di vista dell'umano. Uno dei giochi che ha contribuito al loro successo è senza alcun dubbio *Halo 2* (Bungie 2004), che prenderemo come esempio. In questo testo, diversamente da *F.E.A.R.*, le I.A. dei nemici non eseguono sequenze di azioni dettate da un piano ma una serie di *azioni possibili* dettate dalla loro priorità in quel dato contesto costituito dai tre elementi dello stato mondo, lo stato del giocatore e lo stato proprio. Il livello dell'agire potenziale, dunque, cresce qui ulteriormente rispetto alla precedente potenzialità dei piani in *F.E.A.R.*

Questa priorità, a sua volta, viene determinata non solo da criteri generali comuni a tutti gli agenti e validi nel breve termine ma anche da *desideri* specifici creati per rispecchiare sia una *identità tipologica* (tipi di alieni diversi hanno priorità diverse) che una pianificazione sul lungo termine. Il risultato finale è così un agire adattivo molteplice e variegato che il giocatore *può spiegarsi narrativamente su una base contestuale* e che rispecchia il suo stesso agire e ragionare nel mondo.

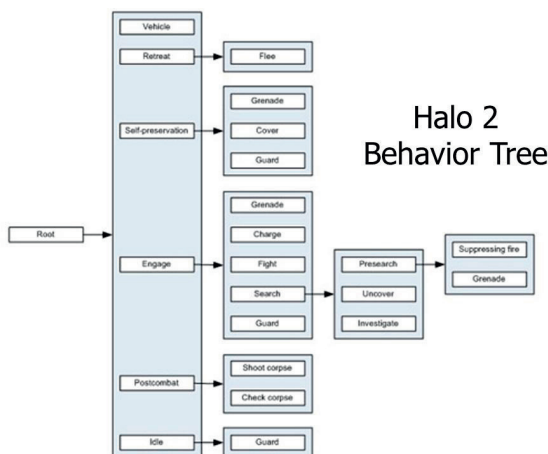


Figura 1. Astrazione del *Behavior Tree* in *Halo 2*⁽¹⁾.

(1) <https://www.gamedeveloper.com/programming/gdc-2005-proceeding-handling-complexity-in-the-i-halo-2-i-ai> (ultima consultazione 23 agosto 2022).

Tanto il sistema G.O.A.P. quanto i B.T. sono dunque sistemi che introducono un minor grado di predeterminazione dell'azione singola in favore di *modelli comportamentali generali e comuni* che sono *contestuali, emergenti* e incentrati su uno *scopo* primario. Questa caratteristica, che ha reso entrambi i giochi molto famosi, è non solo alla base dell'effetto semiotico di una intelligenza superiore rispetto ad altre I.A. ma ha persino dato luogo ad effetti di senso di *cooperazione* fra le I.A. pur non avendo a volte queste nessuna "idea" dell'esistenza di altri agenti ma condividendo semplicemente un dato modello del mondo⁽²⁾.

4.5. *L'agire contestuale efficace*

In ciò che abbiamo appena descritto vi è un aspetto caratteristico dell'intelligenza applicata all'agire molteplice che abbiamo lasciato per lo più implicito: l'esistenza di criteri di preferenza di una certa azione rispetto ad un'altra. Riprendendo l'esempio fatto in chiusura di 4.3, scegliere di colpire da vicino o da lontano l'avversario implica *una diversa attribuzione di un certo valore a entrambe le azioni*. Per il giocatore umano questa preferenza può essere facilmente concepita in termini inferenziali come "se mi avvicino per colpire rischio di essere scoperto e ferito lungo il tragitto, dunque meglio sparare da lontano". Ma come avviene dal punto di vista dell'I.A. questo processo e soprattutto nei casi di sistemi come i B.T? Per spiegarlo riprendiamo qui il concetto di efficacia utilizzato in 3.1 dato che questa preferenza esiste come *entità puramente matematica* per cui *a ogni azione è attribuito un certo valore secondo il contesto in cui avviene* tramite una tecnologia informatica nota come "Utility A.I.". I fattori che influenzano questi valori numerici cambiano chiaramente molto da gioco a gioco, ma ciò che è rilevante per noi è che la *Utility A.I.* fornisce una motivazione di tipo matematico-performativo dietro a ogni azione decisa dall'intelligenza artificiale. Da questo punto di vista, gli NPC esibiscono comportamenti di gran lunga più "performativamente intelligenti"

(2) Al fine di rinforzare questa illusione di collaborazione fra menti diverse viene sovente usato il linguaggio verbale che da un lato ha una funzione pragmatica (aiuta il giocatore a prevedere i nemici) e dall'altro simula una comunicazione fra I.A. che può essere in realtà completamente assente. Per esempio in *F.E.A.R* un soldato può gridare agli altri "ritiratevi!" senza che ciò sia effettivamente una istruzione data da un agente ad un altro.

degli umani che nella maggior parte dei casi non utilizzano un criterio matematico ma piuttosto un sapere di tipo induttivo basato sulla loro esperienza passata.

4.6. *L'agire inferenziale*

Le implementazioni dell'I.A. che abbiamo visto finora sono basate su un adattamento contestuale a ciò che *sta accadendo o è accaduto*. L'intelligenza umana, tuttavia, è anzitutto caratterizzata dall'agire sulla base di ciò che probabilmente *accadrà*, e ciò costruendosi, nel caso del gioco, un *avversario modello*. Una predizione corretta impedirà infatti all'avversario di difendersi, interromperà lo svolgersi di una sua prossima azione o la renderà vana. Sebbene le I.A. della maggior parte dei videogiochi non siano capaci di una tale *scommessa*, esse dimostrano ingannevolmente questa capacità da un punto di vista performativo tramite una tecnica nota come "input reading" usata in giochi molto famosi come la serie dei *Pokemon* (Game Freak) o *Dark Souls* (From Software). Poiché in un videogioco le decisioni devono necessariamente avvenire tramite input, il gioco può leggere questi input prima che il loro effetto abbia inizio e scegliere di conseguenza l'azione migliore per opporvisi. Sebbene questa non sia affatto una previsione (al punto che nella comunità dei videogiocatori viene considerato un vero e proprio *barare*), da parte del giocatore "ingenuo" l'effetto di senso è invece proprio quello di una I.A. che abbia previsto la sua prossima mossa. Questa stessa capacità di lettura del mondo di gioco e dei suoi giocatori è peraltro ciò che consente alle cosiddette "*AI-directors*" di gestire il flusso narrativo di alcuni giochi come *Left 4 Dead* (Valve 2008) o *RimWorld* (Ludeon Studios 2013) manipolando di fatto il comportamento umano tramite la manipolazione di diversi agenti non umani (facendo per esempio comparire un mostro pericoloso in un momento di "calma" in cui tutto sembra andare bene). La capacità di costruirsi un modello verosimile del comportamento umano non è tuttavia del tutto assente nel mondo dei videogiochi. Il suo esempio più notevole può essere visto nella serie di videogiochi automobilistici *Forza* (Playground Games 2005) in cui è stata introdotta, sin dall'inizio nel 2005, una tecnologia chiamata "Drivatar" che *impara* a guidare a partire dal comportamento

del giocatore. Utilizzando una rete neurale di Bayes (BNN)⁽³⁾ l'I.A. registra le azioni e decisioni del giocatore su ogni segmento di una data pista e aggiunge a questi dati sulla sua performatività quelli riguardo il tipo di veicolo.

Grazie a questa operazione denominata “Modellizzazione del Giocatore” (un nome semioticamente significativo), essa diviene capace non solo di emulare il comportamento del giocatore riproducendo verosimilmente quanto egli ha fatto in passato, ma persino di adattare questo modello applicandolo anche su percorsi e veicoli nuovi per il giocatore stesso. Con l'esito finale che l'I.A. prima modella il comportamento intelligente del giocatore (attraverso la raccolta dei suoi dati di gioco) e può finire col sostituirlo trasformando il suo agire umano in un agire macchinico:

Drivatars are the product of a highly complicated system, inspired by an AI technique we call *Player Modelling*. The idea of player modelling is that by gathering enough data of how players play a given game, an AI system can begin to make intelligent decisions that reflect that knowledge. In the case of Drivatar: by understanding how you drive cars in Forza, it can then *drive an AI car in the game as if it was you in the driving seat.*⁽⁴⁾ (corsivo nostro).

4.7. *L'agire esperienziale-mimetico*

L'ultimo baluardo di una specificità dell'agire umano intelligente su una base performativa sembra a questo punto legato alla processualità dietro al momento dell'azione e della decisione: in altre parole all'apprendimento. I videogiocatori umani, infatti, devono tipicamente perdere e fallire numerose volte prima di poter superare con successo una data difficoltà. Più in generale, devono imparare le regole del gioco interagendo con il suo sistema e con altri agenti del mondo anziché imparandole a memoria da un manuale. Ciò molto spesso anche *copiando il comportamento* osservato in precedenza. In questo senso l'apprendimento, quel

(3) E che ha oggi raggiunto una certa notorietà tramite l'implementazione di tecniche di *Deep Learning*.

(4) <https://www.gamedeveloper.com/design/how-forza-s-drivatar-actually-works> (ultima consultazione 23 agosto 2022).

modo di *far tesoro dell'esperienza passata e delle interazioni col nuovo*, è una chiara manifestazione di intelligenza.

Questa capacità di apprendimento è famosa per manifestarsi nel *deep learning* e permette ad alcune intelligenze artificiali come OpenAI di sconfiggere inaspettatamente degli agenti umani anche a giochi estremamente complessi come *Dota2* (Valve 2012). Sarebbe errato tuttavia pensare che solo le nuovissime intelligenze artificiali siano capaci di un agire esperienziale e mimetico. Uno dei primi esempi, infatti, può essere visto in *Dragon Warrior IV* (Chunsoft 1990), in cui l'agire dei compagni del giocatore era determinato da algoritmi di apprendimento per rinforzo. Questi a ogni turno facevano "tentativi" e conservavano nella loro memoria l'esito di questi ultimi al fine di determinare il miglior modo possibile di agire in futuro. Con il risultato finale che l'I.A. risultò fin troppo intelligente⁽⁵⁾, anche in virtù del suo già menzionato accesso matematico al mondo.

È tuttavia nel 2001 che avviene, nel contesto videoludico, una delle più notevoli implementazioni di un agire esperienziale nel gioco strategico *Black and White* (Lionhead 2001). Attraverso un misto di tecnologie algoritmiche viste nelle sezioni precedenti, questo gioco offriva al giocatore la possibilità di far crescere una creatura guidata dall'I.A. sia secondo i desideri del giocatore che in funzione del suo comportamento. Accarezzandola o picchiandola dopo una certa azione, egli le avrebbe insegnato cosa fosse bene fare e cosa no. Ma non solo, la creatura monitorava infatti costantemente il modo di agire del giocatore e avrebbe per esempio assunto un temperamento aggressivo nel caso in cui il giocatore fosse stato frequentemente aggressivo con altri agenti del mondo.

Più recentemente, un gioco al centro dell'attenzione è stato il survival horror *Alien Isolation* (The Creative Assembly 2014), gioco in cui i giocatori hanno avuto *l'impressione* che l'alieno imparasse dai loro comportamenti e si adattasse di conseguenza. In realtà qui la tecnologia in atto è sempre quella degli alberi decisionali, con l'aggiunta tuttavia che questo albero viene *esteso* in funzione delle azioni del giocatore e offre dunque nuove possibilità sempre più efficaci e adatte al contesto. Non di meno, abbiamo ancora qui un adattamento e mutamento dell'I.A. sulla base di esperienze passate.

(5) <https://shmuplations.com/dragonquestiv/> (ultima consultazione 23 agosto 2022).

Infine, non possiamo qui che citare nuovamente il caso di *Forza* che dimostra come le I.A. dei videogiochi siano non solo funzionalmente capaci di imparare dalla propria esperienza, ma persino capaci di agire efficacemente a partire dall'osservazione del comportamento intelligente altrui.

5. Conclusioni

Nel paragrafo precedente abbiamo visto come, alla luce delle diverse tecnologie implementate nei videogiochi prodotti per il grande mercato, sia *funzionalmente* quasi impossibile distinguere l'agire intelligente umano dall'agire intelligente artificiale sulla base del comportamento apparente entro una retorica performativa. E, di conseguenza, come nel contesto videoludico l'agire intelligente umano sia un effetto di senso al pari di quello macchinico⁽⁶⁾. In alcuni casi, abbiamo persino mostrato come, in termini di efficacia, l'agire dell'I.A. sia superiore in efficacia a quello dell'uomo e, entro un'ottica puramente performativa, come ciò porti paradossalmente all'affermazione di una I.A. più intelligente dell'uomo. Un fatto molto caro alle narrazioni videoludiche che proprio su questo costruiscono sovente degli effetti di senso impossibili altrove (Giuliana 2021).

Così la superiorità performativa dell'uomo rispetto all'I.A., che abbiamo mostrato essere una parte del discorso sociale sui videogiochi, si rivela essere in parte il risultato di una *costruzione* voluta dai game designer al fine di garantire una esperienza divertente ed equa: costruzione fondata sulla *castrazione* del vero potenziale performativo di una I.A. iper-avvantaggiata che sa tutto del giocatore umano e che non condivide quasi nessuna delle sue fallibilità (tempi di reazione, input motorio, scommessa, percezione, attenzione, calcoli matematici, etc.). Al punto che la performatività della macchina diventa l'ambizione dell'uomo in contesti competitivi (e-sports) dove delle vittorie dipendono da tempistiche di reazione e di input (pensiamo al *moment 37*) performatate

(6) Al punto che esistono interi giochi come il già citato *Spy Party* basati sulla potenziale ambiguità fra agente umano e agente artificiale. Un ringraziamento al prof. Antonio Santangelo per avermi invitato a sottolineare questo punto.

inaspettatamente dall'umano in un momento storico in cui queste sono credute possibili solo dall'I.A. (Giuliana 2021). Infine, abbiamo mostrato come l'effetto di senso di intelligenza delle I.A. entro un'ottica performativa sia dovuto alla natura semio-narrativa della rappresentazione stessa della performatività (in termini di programmi narrativi, costruzioni di avversari modello, etc.).

Queste considerazioni ci portano a delle conclusioni sulla nostra stessa concezione-costruzione culturale di "intelligenza". Infatti, la conseguenza di quanto abbiamo messo in evidenza è che non solo attraverso i testi videoludici si rinforza l'effetto di senso di una l'intelligenza umana dell'I.A., ma soprattutto che in essi viene messa in crisi la stessa retorica performativa dell'intelligenza *davanti agli occhi del giocatore comune*. Messa in crisi da cui possiamo trarre due conclusioni finali.

La prima è che la causa di questo apparente paradosso di una ambiguità intellettuale fra uomo e macchina è dovuto alla iper-riduzione della complessità del mondo fisico-culturale al mondo virtuale. La caratteristica performativa implicita nella concezione di "intelligenza" può infatti essere valida solo a patto di contestualizzarla entro il quadro semioticamente complesso dell'agire umano incarnato nel mondo e non riducibile alla sua modellizzazione narrativa. Nel momento in cui questo mondo viene invece ridotto a un corridoio pieno di agenti che si devono necessariamente uccidere o a una pista di corse automobilistiche, diventa estremamente semplice modellizzare narrativamente l'agire performativo umano. Similmente, nel momento in cui le interazioni umane vengono ridotte e formalizzate in modo funzionale al gioco, il proprio performativo dell'intelligenza umana viene meno. Alcuni giochi come *Portal 2* (Valve 2011) fanno persino ironia su questa diminuzione dell'umano nel virtuale chiedendo al giocatore di parlare premendo lo spazio sulla tastiera e attribuendo l'incapacità locutiva del giocatore a un danno cerebrale. Questo criterio "ambientale" di complessità interazionale è d'altronde al cuore delle teorie cognitive oggi dominanti e dette "estese" (Newen *et al.* 2018) e che si contrappongono proprio alle modellizzazioni passate incentrate sull'agente e sul suo agire funzionale. Da questo punto di vista, i videogiochi non possono che costituire un'ulteriore prova della validità di una concezione contesto-performativa anziché agente-performativa dell'intelligenza umana.

La seconda considerazione, assolutamente complementare con la prima, è che la riduzione ideologica dell'intelligenza a un criterio performativo è un fatto culturale eticamente problematico e che non ha più veramente motivo di essere accettata. Se infatti dovessimo effettivamente ricreare un esperimento di Turing all'incontrario, non sarebbe probabilmente così difficile distinguere il giocatore umano dall'I.A. Questo proprio perché, a differenza della macchina, il *piacere del gioco* è il primo valore che ha di per sé importanza per l'agente umano. I videogiochi umani, infatti, mettono spesso alla prova i limiti del gioco (Fassone 2017), si divertono tramite il gioco anche a prescindere dal sistema pensato dai suoi creatori, e non smettono di dare senso a quella realtà virtuale nei termini della loro realtà fisica e culturale (pensiamo al tempo speso per creare un certo personaggio!). Così se l'agire umano può dirsi intelligente all'interno della pratica del gioco non è in virtù della sua performatività ma in virtù della *creatività e curiosità* che riflette e della risemantizzazione che permette. Un tema e una differenza al centro di videogiochi come *The Talos Principle* (Croteam 2014).

Di recente, il libro *Ludosemiotica* (2020) di Mattia Thibault ha messo in evidenza questo aspetto e ha criticato la ricorrenza nello studio accademico del gioco di una riduzione di questa attività a una forma di pura competizione in cui l'unico scopo è vincere.

Certamente il gioco performativo (pensiamo per esempio al "META") è un fatto ricorrente nei giochi digitali, ma, in un certo senso, anziché essere caratterizzante dell'umano esso è caratteristico del modo macchinico di comportarsi. Questo a maggior ragione se guardiamo ad alcuni tratti distintivi del "giocatore medio" caratterizzati proprio dalla *non-performatività* manifestata sullo schermo sotto la forma di errori ripetuti senza miglioramenti e persino dell'abbandono: gesti impossibili per le I.A. implementate fino ad adesso nella storia dei videogiochi.

Concludendo, con questo articolo speriamo dunque di aver dimostrato la rilevanza socio-culturale dell'esperienza del videogioco dal punto di vista del mondo in cui essi spingono il giocatore comune a problematizzare il proprio filosofico dell'intelligenza umana sulla base della sua *interazione in prima persona* con una agenzialità artificiale preposta ad affiancarlo ed ad opporvisi. Infine, speriamo anche di aver

dimostrato l'utilità di una prospettiva semiotica applicata all'intelligenza artificiale attraverso la messa in luce degli effetti di senso socio-culturali che nascono dall'interpretazione delle testualità videoludiche.

Riferimenti bibliografici

- BARTHES R. (1964a [2002]), *Elementi di Semiologia*, Einaudi, Torino.
- ECO U. (1979), *Lector in fabula. La cooperazione interpretativa nei testi narrativi*, Bompiani, Milano.
- FASSONE R. (2017), *Every game is an island*, Bloomsbury, London.
- GIULIANA G.T. (2021), "L'Oggettuale e la Persona nella cultura digitale", in: PONZO J., VISSIO G. (a cura di), *Culture delle persona: itinerari di ricerca tra semiotica, filosofia e scienze umane*, Accademia Press, Torino.
- GREIMAS A.J. (1970,) *Du sens: essais sémiotiques*, Volume 1, Seuil, Paris.
- GREIMAS A.J., LANDOWSKI E. (1976), « Analyse sémiotique d'un discours juridique » in Greimas A.J. *Sémiotique et sciences sociales*, Paris, Seuil.
- NEWMAN A., DE BRUIN L., GALLAGHER S. (2018), *The Oxford Handbook of 4E Cognition*, Oxford University Press, Oxford.
- LOTMAN J.M. (2022), *L'asimmetria e il dialogo nelle strutture pensanti*, La Nave di Teseo, Milano.
- SKINNER G., WALMSLEY T. (2019), "Artificial Intelligence and Deep Learning in Video Games A Brief Review", 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 404-408, doi: 10.1109/CCOMS.2019.8821783.
- THIBAUT, M. (2020) *Ludosemiotica. Il gioco tra segni, testi, pratiche e discorsi.*, Aracne, Roma.
- TURING A.M. (1950), *I.—COMPUTING MACHINERY AND INTELLIGENCE*, *Mind*, Volume LIX, Issue 236, October 1950, Pages 433-460, <https://doi.org/10.1093/mind/LIX.236.433>.
- VOLLI, U. (2005) *Manuale di semiotica*, Laterza, Roma.
- WEXLER J. (2002), *Artificial Intelligence in Games: A look at the smarts behind Lionhead Studio's "Black and White" and where it can and will go in the future.* <https://www.cs.rochester.edu/~brown/242/assts/termprojs/games.pdf>.

Sitografia

<https://www.youtube.com/c/AIGamesSeries>

<https://www.youtube.com/c/MarkBrownGMT>

<https://gamerant.com/best-games-advanced-ai/>

<https://www.thegamer.com/best-enemy-ai-video-games/>

<https://www.rockpapershotgun.com/7-games-with-the-best-ai>

<https://dataconomy.com/2022/04/artificial-intelligence-games/>

<https://www.cheatcc.com/extra/top10gameswithstupidai.html>

<https://whatculture.com/gaming/10-video-games-with-the-worst-enemy-ai>

<https://shmuplations.com/dragonquestiv/>

<https://www.gamedeveloper.com/programming/gdc-2005-proceeding-handling-complexity-in-the-i-halo-2-i-ai>

<https://www.gamedeveloper.com/programming/gdc-2005-proceeding-handling-complexity-in-the-i-halo-2-i-ai>

NARRARE L'IA

FIT DIFFERENT! AN EXPLORATION OF APPLE FITNESS+ ARTIFICIAL AFFECTS MACHINE

ELSA SORO*

TITOLO IN ITALIANO: Mettiti in forma diversamente! La macchina degli affetti artificiali di Apple Fitness+

ABSTRACT: la crescente domanda e offerta di servizi e soluzioni digitali per il fitness, incrementatasi durante la pandemia di Covid 19, ha presumibilmente riarticolato le sfere dello sport e del tempo libero, insieme al loro significato sociale e culturale. Questo lavoro intende indagare gli effetti di senso derivati dalla nuova valorizzazione della pratica del fitness. In particolare lo studio si interessa a come le tecnologie basate sull'intelligenza artificiale ambiscano a sostituire la motivazione infusa da attori umani negli spazi adibiti al fitness. Prendendo in considerazione l'ambito di studi ormai consolidato che si concentra sulle dimensioni socioculturali del fitness digitale, questo articolo intende partire da come le app di fitness, in interazione con il corpo dell'utente, risemantizzano le motivazioni e le aspirazioni verso questo ambito di esperienza.

L'articolo analizza le interfacce, le *affordance* e le narrazioni visive prodotte Fitness+, l'app per il fitness di Apple Inc., e confronta quest'ultima con un campione di altre app per il fitness basate sull'intelligenza artificiale. Si utilizzerà una lente semiotica per interpretare il modo in cui il corpo, i movimenti e gli affetti sono inseriti nella cornice dei valori di "progresso" e "diversità" espressi dal marketing dalla *big tech* californiana.

KEY-WORDS: digital fitness, Artificial Intelligence, semiotics, affects theories

1. Stay (fit and motivated) at home

Since the advent of the Covid 19 pandemic, due to the limitations imposed on conventional sports activities, the use of digital sport and fitness services has grown significantly. These digital solutions include a

* Università degli Studi di Torino.

vast array of technological services and applications, ranging from on-demand online training to free workouts on YouTube and TikTok, etc. At the same time, the market for wearable devices such as smartwatches for motion tracking has multiplied its revenues.

The increasing demand for and offer of devices for digital fitness services and solutions has arguably rearticulated the fields of sport and leisure, along with their social and cultural meanings. Indeed, sport and fitness facilities have been relocated to domestic or public spaces and the presence of fitness professionals and peers has been replaced by digital personas who act as trainers and online communities. In this regard, fitness services have proposed different solutions to digitally recreate the motivation for sport, whether it be competing in a race or sharing the effort and energy with peers. Seemingly, online fitness communities share motivational hashtags and images of achievement.

Arguably such processes of remediation and relocation of workouts and physical exercises have ultimately impacted the body and its own proprioception. If, generally speaking, in the traditional fitness facilities individuals imitate the movement of the instructors while looking at their reflections in the classroom mirror, the use of online applications with their digital screens transforms this awareness of movement and coordination. Furthermore, the infusion of motivation and the monitoring of exercises which, along with the supplying of equipment, can be considered at the core of the services delivered by the gymnasiums with their staff of instructors, has had to be digitally recreated and therefore discursively readapted.

In this regard, the use of artificial intelligence in digital fitness (Farrokhi *et. al.* 2021) has been implemented by fitness service providers to fill the gap left by the absence of a shared physical space where bodies and movements were at the same time monitored, compared and therefore motivated to improve, as well as the absence of person-to-person interaction that was supposed to create engagement towards fitness.

Like other fields and industries, the premise of the AI functioning in fitness is the availability of big amounts of data from wearables, sensors, cameras, and other connected devices related to eating habits, workout performance, and fitness goals. According to the marketing discourse conveyed by the apps (Meng 2021), the processing

of such data can generate AI-powered recommendations to optimize user performance.

The motivation factor, traditionally supplied by a human trainer in sport facilities, has been recreated in many apps and services through the visualization of live performances and their metrics thanks to bio-data tracking. The screens of smartphones or smartwatches have thus become an extension of the internal organs⁽¹⁾ that can be *seen working* on the screens and contrasted with previous performances by other users. This way, the physical sensation during the workout can be transformed into a number and metric to which a value can be attributed. In turn, those data become analytics that feed the creation of new metrics and therefore predictive data.

The possibility of tracking motion through sensors and AI-powered computer vision can also ensure that users perform the exercises *correctly*, and at the same time produce a new archive of digital images of the human body on the move, aggregated with metrics and values.

2. “Fit” different: Apple Fitness+ motivator

Within this ideology of constant improvement and robotization of the human body that can be visualized on digital screens as data and metrics, the different services on the market have to differentiate themselves with a variety of commercial offers. For each of them the visual restitution of the motivation to exercise through technological solutions has to replace human and spatial interaction as experienced in physical fitness facilities.

Through the commercialization of its first wearable device that can carry out electrocardiograms (ECGs), the Apple watch and the consequent launch of Apple Fitness, the well-known Californian manufacturer of personal computers and software has become one of the major digital fitness providers.

The launch of a fitness app by one of the biggest tech companies in the world can be considered a very relevant case within the ongoing digital reshaping of the leisure spheres and of body experience. In order

(1) Carbone et al. have extensively problematized the relations between screens and the human body within what they call an *anthropology of screen experiences* (see particularly Carbone et al. 2021).

to grasp certain dimensions of this transformation, this piece of work will focus on the analysis of the Apple Fitness ecosystem of values and aesthetics within which fitness experience is included.

Through the examination of Fitness+ affordances, interfaces and visual textualities this paper will articulate an investigation into how the biometric tracking embedded in the digital fitness providers is visually pertinentized (or not) and which effects of meaning such (un)visibilization strategies convey.

The data collection process is the result of a netnography carried out by the author as a user of the service over more than 5 months, during which she joined the Fitness+ “club” and practiced a variety of workouts as part of the offer. Such empirical observations have been combined with desk analytics and a descriptive analysis of the service interface and visual contents.

Regarding the structure, this study is articulated as follows: the first part will look back in time to a famous case of virtual fitness in the 1980s featuring Jane Fonda and its legacy for fitness influencers on social media; the main strands of research on digital fitness and how they dealt particularly with the gendered and normative body of female users will be reviewed in the second part; then a sample of fitness app commercialized as “AI based” will be reviewed. For each of them the opposition human–machine as displayed by the app visual and textual discourse will be unmasked. Such sample of apps will be functional to introduce the “difference” and “diversity” represented by Fitness+.

Through an analytic description of discursive and enunciative arrangements of Fitness+ in conjunction with the user body accessorized with the wearable Apple devices, the work will ultimately attempt to shed light on the integration of bodily movement and exercise experience in the wider affective and aesthetic system produced by tech corporations such as Apple.

3. Jane Fonda on TikTok

Despite the acclaimed success of digital fitness in concomitance with the spread of the global pandemic, the act of training with virtual

instructors is not a novelty. In the 1980s Hollywood star Jane Fonda popularized the format of virtual classes with her *Workout Starring Jane Fonda*, a series of recorded classes supported by videotape. The initiative stemmed from the actress' intention to subvert the traditional masculine dominance in gyms and create a space where women could "burn calories" and build muscles. The workouts were shot in a theatrical set, in a ballet studio located in Beverly Hills where the actress led a group of fitness practitioners.

After its launch, the product was the top-selling VHS tape for a few years, "positioning" the actress as an influencer *ante-litteram*. More than 30 decades later, Jane Fonda's workouts have appeared on YouTube and Tik Tok, revamping her characteristic leg warmers and striped leotards and pelvic lifts.

These days on social media one would tag Jane Fonda's workouts as "Fitspiration" or "fitspo" (a mix of the words "fitness" and "inspiration"), a popular hashtag that circulates in social media and refers to a "global philosophy of health, strength and empowerment" (Tiggemann & Zaccardo, 2018). With the emergence of the social media platforms of video sharing such as Youtube, Instagram or TikTok and the consequent phenomenon of the influencers, the free offer of contents related to fitness has radically increased and consequently the "authority" of the trainers has started to be "distributed" among a plethora of actors whose credibility may be measured by quantifying the hits and likes that they receive. This phenomenon reverses the model launched by Jane Fonda where arguably the role of celebrity preceded that of fitness influencer.

When the global pandemic started in 2020 with a series of national lockdowns and corresponding restrictions, the consumption of mediated fitness contents drastically escalated. As a response to the temporary closing of fitness facilities, the home was refunctionalized as a place of "everything", including workouts. In this vein, several videos on social media platforms instructed users on how to "use household items" such as plastic bottles of water or packets of rice or even books as dumbbells.

Along with the diffusion of free fitness-related contents on social media platforms with the emergence of innumerable influencers, a variety of mobile applications have appeared on the market. These services

can be differentiated between the workout apps that show exercises and explain how to do them and the tracking apps that track movement (either workout or normal movement).

Both typologies can be paired with wearable devices that collect and visualize the performance data in real time (typically the heartbeats). Furthermore, numerous health- and fitness-related apps include gamification features in the attempt to engage users by rendering routine aspects of everyday life into games; these are thus referred to as “serious games” in the medical literature.

If the market has certainly noticed the expansion of the demand for digital fitness and rapidly reoriented business models to this digital transformation, in Academia a nascent literature has problematized the psychological, social and cultural aspects of online fitness cultures. In the following section, some relevant strands of research will be examined as background to this paper.

4. Body of researches on Fitspiration: a quick review

In their bibliometric study of this field to better understand the current state of research, Liu and Avello (2021) found that the contributions to the topic of digital fitness belonged primarily to the fields of medicine, computer science, and health sciences. Nonetheless, the sociocultural dimensions of digital fitness have also been addressed by scholars, with a special focus on conceptions of body appearance and paying particular attention to the equality of treatment between male and female bodies (Fardouly et al. 2018).

In this regard, scholars of sport have used affects theory to explore the ways in which women’s exercising “bodies intersect with embodied feelings and affects, sexual subjectivities, and desires” (Roy 2013, 331; also see Pavlidis and Fullagar 2015; Waitt 2014).

Drawing upon feminist theorizations of affect and the body (Ahmed 2004), Kim Toffoletti & Thorp analyzed the BBG fitness community on Instagram and focused on the emotions and sensations that reverberate through bodies as they circulate and are amplified via the interactive and content-sharing features of Instagram; they support the

thesis that mediated fitness culture “complicates linear associations between fitspiration media and women’s body image concerns” (p.2).

A post-feminist biopedagogical perspective (Wright, 2009) guided Camacho et al. (2019) in their analysis of how young females negotiate discourses around health and fitness whilst engaging with Instagram. The concept of perfection entangled with the neoliberal ideologies of continuous self-improvement is at the center of Camacho and Gray’s (2021) examination of the social media presence of Patry Jordán, a famous Spanish fitness influencer followed by young women. A further examination of young female bodies in relation with health-related social media by Goodyear and al. (2021) investigated the media’s influence in the learning processes associated with a normative healthy body.

A complementary subject of research within e-fitness and e-health have been the data assemblages produced by the self-tracking activities enabled by body sensors and self-tracking devices such as the Apple Watch or Fitbit. Australian researcher Deborah Lupton (2018) considers the data generated by self-tracking devices to be “lively”, since they are:

generated from life itself (...), open to constant repurposing by a range of actors and agencies (...). Due to the advent of algorithmic authority and predictive analytics that use digital self-tracked data to make inferences and decisions about individuals and social groups. (2018, p.14)

In her book “Quantified self”, a name borrowed from QS, “an international community of users and makers of self-tracking tools who share an interest in “self-knowledge through numbers”, Lupton describes where the harvested wearer data go and for what purpose by visualizing privacy concerns. This use of the lively data by actors and agencies and third parties within the capitalist system is considered a biopolitical goal of governments that encourage healthy behaviours and “promote a culture of dataveillance” (Lupton 2016).

This body of studies about the so-called “fitspiration” media practices has mainly focused on the exposition, often gendered, or the surveillance of bodies in social media through empirical research on users’ experiences of self-tracking, or on the use of the app and services

including strategies that practitioners utilise “to avoid conforming to a form of ‘techno-optimism’ (Esmonde, 2018)”. Less attention has been given to the construction of the fitspiration experience by the service itself and application affordances, interfaces and visual discourses showcased on the device’s screen. In this regard, this piece of work aims to start from the fitness app and the discousivization of the assemblage with the user body. In particular how the interaction between the human body-machine is problematized by the commercial discouse of the fitness apps.

Such considerations will then pay the way to present the value proposition of Fitness+, consisting in “being different” from the rest of the AI based fitness apps.

5. Artificial intelligence *vs* human body in the “rest of” fitness apps marketing strategy

In the vast territory of fitness apps, we selected a sample of applications sorted among the first five in the Apple Store and in the Google Store (during the period December 1st — May 1st 2022), by using the keywords “Fitness” “App” and “Artificial Intelligence”. The small series of apps appearing among the first five entries in the two stored consisted of FitnessAI, Freeletics, Fitbod, Aaptiv Coach, Gymfitty. Of the five apps, the corresponding landing page of website where the product is described and marketed by the seller, will be analysed in order to grasp the different valorization of the AI — in interaction with the human body.

The first app of the sample uses the Artificial Intelligence already in the name, FitnessAI, followed by the slogan “Get Stronger with A.I.”. It is an app for iPhone that uses artificial intelligence to generate personalized workouts. It is based on 5.9M workouts for which the AI optimizes sets, reps and weight for each exercise the user works out. According to the seller, the app would *outperform* any human personal trainer. Accordingly, FitnessAI website feature image consists of a dark background out of which a human silhouette whose facial features are barely visible is lifting a workout bar. This occultation of the human face could reinforces the emphasis of the *power* of the technology over

the human. This is also reinforced by the fact that the slogan — “See the Algorithm at Work” — provides the technologies of an agency.

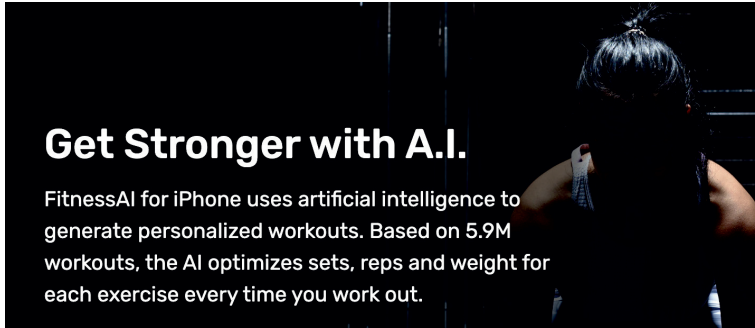


Figure 1. FitnessAI app.

The header of Freeletics website consists of a male human figure caught while performing a plank position. His gaze, stars at the spectator, as a form of *embrayage*. In the scene no technological device appears and the slogan “To the endorphin lovers: bodyweight, weights, running” stances a sort of organic (and therefore anti-technological) isotopy. In the app discourse, the occultation of any technologic element in



Figure 2. Freeletics app.

favour of the mere human body exposed with its organic components is further reinforced by the app description below the website header: “The AI does the thinking, you do the sweating”. In this case, the AI seems to be ancillary of the human one and of the human and organic body which in this case seems to hold the agency.

In the spectrum of different types of marketing discourses displaying a technological–machinic valorization *vs* a human–bodily and organic one, FitnessAI and Freeletics represent the two more polarized positions. In between, the other three apps represent more “moderated” marketing choices that lean towards the interaction technology–humans.

In such vein, Fitbod website heater presents the actual app visualized in a cellular phone.

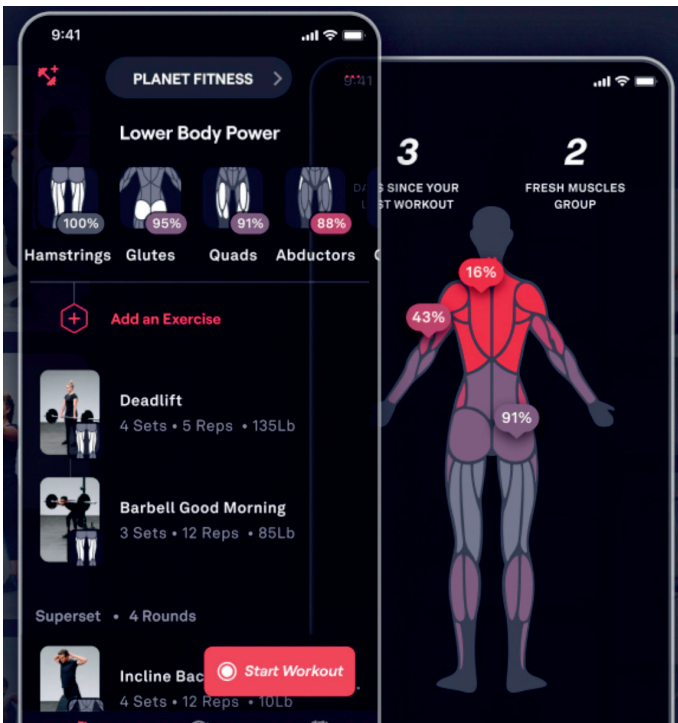


Figure 3. Fitbod app

The app shows a gallery of digital images of muscles targeted by different exercises featured by the app: “Built for better. Workouts that

improve as you do”, the slogan says. Fitbod value proposition is indeed based on a learning-together mechanism. According to the seller, Fitbod creates workouts tailored to the user based on muscle usage, exercise achievements and track the user fitness progress. The routine created by the app then change over time and by algorithms and machine learning (ML) modify training volume based on previous exercises so to maximize outcomes. According to such offering, machine and humans progress together.

A similar interaction human-machine is thematized by the Gymfitty landing page. Here the heater is in motion: a video of different human actors, a young woman boxing, another jumping a rope, a young man lifting weights, another doing push up. All have extremely muscled and fit bodies and a close-up on them focuses the tone of their sweating fleshes. Though, at the bottom of the page a image of a runner positioned at the race starting point has at his left a chat box where a conversation with a trainer bot is going on: “Ready”, “Yes”, “Start running”.

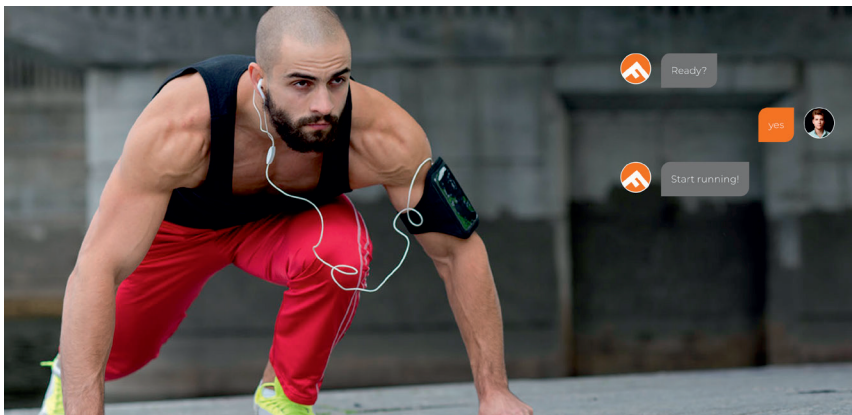


Figure 4. Gymfitty app.

According to the seller, this app has developed complex algorithms that use artificial intelligence for creating personalized sessions of work. The app monitors the performance in real-time and adapts the workout accordingly. The description of the app points out: “just as if you have your trainer right there alongside you”. In the exact moment of the performance, that right here-right now, the human body does not

act alone. Arguably such close-up of the heater stands for the external machinic eye of the AI who constantly monitor “you”, in the very moment of the physical performance.

The visual architecture of Aaptiv Coach website is conceptually different from the previous examples. The actual app is displayed in the heater that showcases its different functionalities. From the functioning point of view, Aaptiv Coach provides Aaptiv members with a personalized coaching plan that incorporates fitness, mindfulness, and healthy habits, as part of a holistic approach to maintain a *healthy lifestyle*. In the section of the website “How it works”, the user is invited to “Enter your goals”, “Choose a workout based on duration, trainer, music, and more”, “Put on your headphones and let our trainers guide you”.

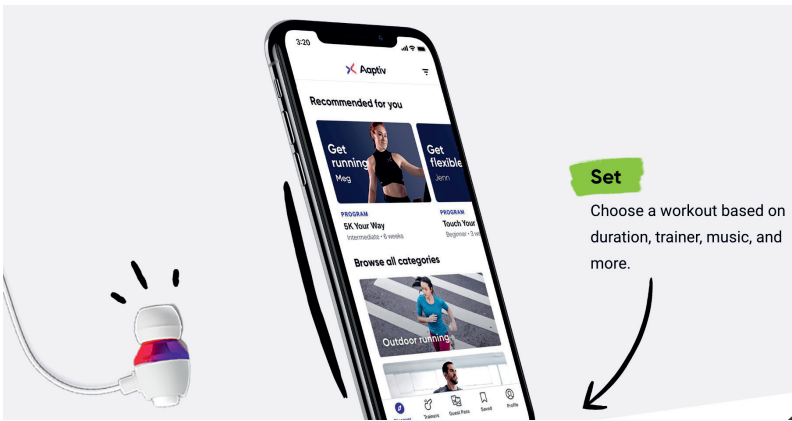


Figure 5. Aaptiv Coach app.

All the steps are figurativized by a mobile phone displaying the correspondent section of the app. On each side, an indexical arrow points at the bottom to push.

According to the app mission, not only the service assists the user while working out. Also, the app offers *lifestyle* programs based on the user particular objectives, current exercise levels, eating habits. All those data are gathered through wearable devices such as a wristwatch or activity tracker. In this way, thanks to the presence of a wearable device, Aaptiv Coach promotes a permanent interaction between the body and

the machine that goes beyond the mere exercise and seems to spread more deeply into the user lifestyle.

In the following section this study will approach how such continuous and pervasive feedback between machine and human is packed into the Apple Inc. ecosystem and frame of values.

6. “It is amazing what you can do with your wrist”

Since the 1990s *Think Different* campaign, Apple’s communication strategy and visual identity have been widely studied. Semioticians are not an exception in this regard. In such vein, Jean Marie Floch in *Identités visuelles* (1995) analyzed the “mythogram” of the curvy bitten apple, in comparison with the IBM one of which it constitutes an inversion. In opposition to “competency” as the value of the IBM logo, the bitten apple represents the embodiment of “creativity”, “conviviality” and “freedom”.

This “alternative” character of the Apple brand within the IT industry is further marked by the Switch advertising campaign launched by Apple on June 10, 2002. According to Apple, “The Switcher” refers to a person who changes from using the Microsoft Windows platform to the Mac. The eccentric figure of Apple co-founder Steve Jobs and his premature death have contributed to the fact that every Apple product enters the mythical sphere. The construction of the myth has also been contributed to by the popular Apple events in which the company unveils new products and services that have previously been kept secret to create a buzz.

At one of these launching events, in September 2014, the Apple CEO announced Activity, an exercise tracking companion app. The ceremony was premiered by the presentation of the iPhone 6, after which, at Tim Cook’s words “One more thing”, the first images of the Apple Watch were shown. “It is amazing what you can do with your wrist”, Cook said, connecting the Apple products and devices with an idea of lifestyle based on the “stay active” motto.

In fact, along with the launch of the first models of the Apple Watch, two applications had been developed: a Fitness app to monitor all activities throughout the day and a workout app with which it was possible

to set specific goals for specific types of workout.

The promotional video of the campaign streamed during the launch ceremony described a “typical” daily user routine. The plurality of uses that the device incorporates were figurativized by different actors engaging in different routines: an elegant businesswoman busy with work meetings, a muscular man performing strength workouts, a cyclist involved in performative riding in a natural environment and two athletic women running outdoors in a striking rocky coastal landscape. Despite their different narrative programs, each of them succeeded in making it through the day and getting where they wanted to.

This sense of accomplishment in both phones and watches is figurativized by the “activity ring”, the Apple interface through which users who wear the Apple watch know whether they reach the goals which they have set themselves (or which the app has suggested on the basis of the data introduced by them). This figure is composed of 3 concentric circles, a red one that shows how many active calories the user has burned; a green one that displays how many minutes of brisk activity the user has performed and a light blue one that shows how many times during the day the user has stood and moved for at least one minute per hour. The goal of a “good” user is to close all three rings by the end of the day, with the rings resetting the next day so people can close them all over again. Moreover, along with the daily challenge, the users can set weekly or monthly ones and win “badges”.

Since the launch of the Apple Watch, the Californian computer manufacturing company has not only expanded and differentiated its array of devices: at the same time, having — as the CEO enthusiastically announced — literally the control of the user’s wrist, another sphere of life has been assimilated into the Apple ecosystem (and by so doing another series of user data concerning the body) with the consequent possibility of carrying out forms of control over it.

7. Close your ring!

Since 2014, the wristlet has expanded its capacity of doing things with the wrist to continue with the expression used by Cook. On September

2020 the Apple Special Event was celebrated without an audience because of the restrictions due to the Covid 19 pandemic. This time Cook was shown while walking in the circular groundscraper of Apple Inc.'s corporate headquarters in Cupertino. In his speech, Cook made a reference to the challenge represented by the pandemic and how incredibly people had adapted. He then emphasised the role of Apple products in overcoming isolation, bringing people together and moving forward.

The health isotopy that emerged in his speech in connection with the pandemic was then reevoked with regard to the Apple Watch's new features, such as the heart monitor which, according to the words of Apple's CEO, "in certain circumstances saved people lives".

On this occasion Cook gave the floor to Jay Blahnik, a former Nike contract athlete and Nike consultant, in his capacity as Director of the new area Fitness Health Technologies, in the meantime implemented for the introduction to the public of Apple Fitness+. While launching the service, Blahnik was located in a set for fitness. Behind him, in the background, a graffiti reproduced the three colours of the rings. The Director presented the features in detail, for instance the metrics, showing in real time on the user's wrist and on the screen of the iPhone all the types of workouts available. He emphasized the plurality of possible locations (at home, at the gym, outdoors, in a hotel room) where the workouts could be performed; the variety of workouts and the curated playlist were two other features promoted by Blahnik.

A promotional video of Fitness+ presented the community-focused dimension of the service. In this video, a voice off affirms "Welcome to the club, the club has no rules", while different environments where people exercise pop up ranging from a fancy house, tiny flats, parks, urban areas surrounded by blocks, mountain peaks, garages, etc. The voice off continues: "The club has no door and no roof, there are walls that can be thrown down". The selection of people who exercise accurately displays a variety of diversities in terms of gender, ethnicity, disabilities, and even non-human actors, such as pets. The voice off states at the end: "the club is the largest club in the world because the club is the world".

The imaginary of the community regarding digital fitness is surely not a novelty. In fact, one of Fitness+'s main competitors (with 1.3 million

subscribers), the paid fitness service Peloton, derives its name from the French word “platoon”, which means a group of riders — usually in a race — who save energy by riding closely together. The semantic of community, from the Latin word *communitas* implies the sharing of a set of norms and expectations regarding different spheres such as the use of a language⁽²⁾. The philosophical complexities of the notion will not be the object of further inquiry here, but a reference to the work of Roberto Esposito on the opposition *Communitas* vs *Immunitas* (2020, 2022) brings out the implication of the immunization occurring in a community against any foreign element that appears to threaten them from the outside.

From a semantic standpoint, a club is not exactly a community. Indeed, according to the Oxford Dictionary, the first entry for “club” regards the sphere of sport: “an organization for people who share an interest or do a sport or activity together”, followed by “an association of persons for some common object usually jointly supported and meeting periodically”.

Closer than a community, the club therefore entails a form of membership that gives access to certain private facilities (ranging from sport to nightlife). However, in contrast to this semantic, as declared by the promotional video, the Apple fitness club has no rules. In the oxymoronic notion of a club without doors, emphasised by Blahnik’s affirmation that “everyone feels invited to the party”, lies the hybridization of exclusivity and inclusivity that characterizes the whole body, aesthetics and discursive experience of Fitness+.

7.1 *The Club*

Along with the notion and the correspondent spatiality of a “club”, Blahnik referred to the Fitness+ service as a “studio workout”. According to the Oxford dictionary, a studio is firstly “a place where movies are made or produced”. The Director of Apple Fitness continued: “We built the studio in a way that would allow shooting all the angles to make the right choices to show just the right angle at just the right time”.

(2) Translation of the author of the definition of the Italian Word “comunità” according to Treccani dictionary.

In Apple's new 23,000-square-foot space, in a wood-panelled and wood-floored fitness studio with three sliding glass doors opened to reveal a verdant garden, the classes are streamed in cinematic spatiality thanks to the high-end Super 35-format cinema cameras mounted on robotic arms that enable smooth movement, in high definition.

The plastic transformations of chromaticity and luminosity of the space afforded by the mobile cameras articulate the production of a mobile and dynamic viewpoint. The multiplicity of angles — seven robotic cameras focus on the trainers from the front, sides, and even directly above — accentuate different points of attention: not only are the body and the movement of particular muscles highlighted, but also sometimes at the centre of the visual composition is the face of the trainer, smiling despite the effort or singing along to a track from the selected playlist. Occasionally the trainer invites the user to check the ring on their wrist so that the focus moves out of the enunciative space and onto the user's body.

There are usually three trainers per workout. The leader is a specialist in the workout, and there is at least one modifier. By having “delegates” of different kinds of potential users on stage, the narration underlines this dimension of the club's inclusivity. The same effect is produced by the offers of workouts that feature pregnant women.

7.2. The diversity fair

Where other apps use animated characters or AI bots, Fitness+ features a series of trainers, diverse in terms of gender, ethnicity, and background, ranging in age from their 20s to their 60s. Browsing the profiles on the Fitness+ website, the trainers seem a mosaic of tokens of an accurately selected diversity within American multiculturalism, aligned with the values of Apple workforce culture.

The diversity is not only ethnic, with members of Latino, Afro-American and Afro-Caribbean origin, but also linguistic: both American and British English are used as well as Spanish, sign language and gender neutral language. Furthermore, the trainers figurativize a repertoire of a-normative bodies, challenging the “normal” fit-looking appearance and also including bodies with surgical prostheses, as in the case of Amir who has a prosthetic leg.

In the profile gallery, a little information about each trainer is provided below black and white profile pictures of the smiling trainer, viewed both frontally and from the side. In the profiles, not only are the corresponding sports discipline speciality and professional CV (i.e. former UCLA football player) presented, but also contextual episodes from the trainer's life are displayed that have apparently nothing to do with the sports sphere ("after losing his left leg in an accident", "a busy mother", "she went from playing bass guitar in a band to being featured in fitness magazines", "Jessica has surfed in 80 countries around the globe and she is unapologetically obsessed with anything in leopard print"). Such "additional" information contributes to expanding Fitness+ beyond its own borders and turning it into a wider lifestyle Destinant.

This role is also reinforced by the expansion of the app into social media, as conveyed by the presence of the Instagram Account at the bottom of each trainer profile on the Apple website. If one browses a trainer's social media profile, a relevant degree of consistency will be found with the persona showcased within the fitness app. Every trainer's IG profile refers to their involvement in the Fitness+ project, along with motivation quotes, tastes, beliefs (i.e. "gay, black and proud"; "proud latina woman"; "social justice advocate").

By surfing the posts and monitoring the stories, one can observe that such tastes and beliefs are narrated throughout the digital self-presentation of the trainer, who is often portrayed while involved in activities in sports facilities or beautiful outdoor locations or engaged in leisure activities (such as music festivals). We can refer to this as a reverberation of Apple Fitness values with a further expansion of its ecosystem beyond its own digital and commercial borders.

8. Fitness and affects

In the description and analysis of the "Fitness+ experience" textualized in the Fitness app and related social media platforms, we have noticed the magnification of what we can call an *affective dimension* of digital fitness and consequently a narcotization of its merely quantifiable and metric aspects.

Even if this latter sphere is present in the affordances, interfaces and visual discourses carried out by the trainers, the analytic and empirical observations of the app lead to a consideration of the predominance of the affective dimension. The ring, to which the measuring function is delegated, arguably denies this numeric quantification function. Instead, through the circular number-free representation of the “work done”, the ring visual rhetoric seems to obscure precise measurement in favour of a more impressionistic boosting of an aesthetic lifestyle.

Through both technical and discursive and figurative arrangements, without forgetting the role of the music selection, the user has been provided with a high quality spectatorship experience at the same time as he is training. The user *watches* the show of the workout performed by the three actors while listening to the stories and anecdotal experiences and even jokes being told and to the music selection accurately chosen and commented.

Differently from other fitness services, where the value proposition has relied on the AI’s capacity of monitoring movement in real time and recommending certain moves, Fitness+ appeals to different motivations for engaging in digital fitness. The access to a fitness “club” populated by diverse and multicultural subjectivities — the progressive and multicultural West Coast in a nutshell — has to do, in our view, with the articulation of a promise of belonging.

This aspirational status, promised by the visual and discursive narratives of Fitness+, can be materialized through the possession and use of different Apple devices that allow the individual to be more (or less) integrated in the Apple ecosystem. The feature of the syncing between Apple devices with which the Mac iPhone, iPad, Pod and Apple Watch all transfer data and are reciprocally kept up to date can be considered extendible to their human wearers. In an assemblage, human and non-human actors are kept synchronized and aligned with Apple in an aesthetic lifestyle prescription system. In order to critically describe this process of agency co-creation, we have dared to create the expression Artificial Affects in order to refer to these “affective fabrics” (Kuntsman, 2012) of Apple lifestyle culture, where the “feeling” good supersedes the biometric “intelligent” quantified experience.

9. Conclusive notes

At the moment of writing, Apple has just announced the acquisition of AI Music, a start-up that uses artificial intelligence to generate personalized soundtracks and adaptive music. The implementation of this technology would generate dynamic soundtracks based for instance on the user's heartbeat, adjusting to workout intensity. In our view, this does nothing but amplify the "artificial affects" frame by further synchronizing the user "feeling" with the Apple aesthetical ecosystem.

This piece of work stems from the phenomenon of relocation of an entire sector of leisure such as fitness in the digital space, the gap of meaning that such a displacement has produced and how it has been filled by the AI technologies applied to digital fitness. In examining the Fitness+ "solution" in comparison with other AI based fitness apps, the article has attempted to describe the "place" of biometrics and that of affects at the heart of the Apple ecosystem. From the observation and analysis of device affordances, interface and visual discourse it seems clear that they should be comprehended in their combination with the user body and senses. Far from being exhaustive this paper's intention has been to sketch an approach through which to engage with the digital affects, aesthetics and lifestyles factory put in place by tech capitalism that has a long way to go.

References

- AHMED S. (2004), *The Cultural Politics of Emotion*. New York/Abingdon: Routledge
- BAL M. E. DÓCI (2018), *Neoliberal ideology in work and organizational psychology*, "European Journal of Work and Organizational Psychology", 27:5, 536-548
- CAMACHO-MIÑANO M.J., MACISAAC S., E. RICH (2019), *Postfeminist biopedagogies of Instagram: young women learning about bodies, health and fitness*, "Sport, Education and Society", 24:6, 651-664
- CAMACHO-MIÑANO M.J., S. GRAY (2021), *Pedagogies of perfection in the post-feminist digital age: young women's negotiations of health and fitness on social media*, "Journal of Gender Studies", 30:6, 725-736

- CARBONE M., LINGUA G., S. DE SANTIS (2021), *Making Screens Functions and Technical Objects*, "Screen Bodies", 6:2, 1–22
- DERNDORFER J., DISSLBACHER F., LECHINGER V., MADER K., E. SIX (2021), *Home, sweet home? The impact of working from home on the division of unpaid work during the Covid-19 lockdown*. "PLOS ONE" 16(11)
- ESMONDE K. (2018), *Tracing the feedback loop: a Foucauldian and actor-network-theory examination of heart rate monitors in a physical education classroom*, "Sport, Education and Society", 1–13
- ESPOSITO R. (2022), *Communitas. Origine e destino della comunità*, Torino: Einaudi.
- (2020), *Immunitas. Protezione e negazione della vita*, Einaudi, Torino.
- FARDOULY J., WILLBURGER B.K., L.R. VARTANIAN (2018), *Instagram use and young women's body image concerns and self-objectification: Testing mediational pathways*, "New Media & Society" 20:4, 1380–1395.
- FARROKHI A., FARAHBAKHS R., REZAZADEH J., R. MINERVA (2021), *Application of Internet of Things and artificial intelligence for smart fitness: A survey*, "Computer Networks", 189: 1389–1286.
- FLOCH J.M. (1995) *Identités visuelles*. PUF: Paris.
- KUNTSMAN, A. (2012) "Affective Fabrics of Digital Cultures," in Karatzogianni A. and Kutsman (Eds), *Digital Cultures and the Politics of Emotion* Palgrave Macmillan, 1–17.
- GOODYEAR V., ARMOUR K.M., H. WOOD (2019), *Young people and their engagement with health-related social media: New perspectives*, "Sport, Education and Society", 24:7 673–688.
- LIU Y., M. AVELLO (2021), *Status of the research in fitness apps: A bibliometric analysis*, "Telematics and Informatics", 57.
- LUPTON D. (2018) *Digital Health. Critical and Cross-Disciplinary Perspectives*, London and New York, Routledge.
- (2016), *The quantified self: A sociology of self-tracking*. Cambridge: Polity Press.
- ROY G. (2013), *Women in Wetsuits: Revolting Bodies in Lesbian Surf Culture*, "Journal of Lesbian Studies", 17: 3–4, 329–343.
- XIANGKUN M. (2021), *Fitness app Application Research Based on big data and algorithm*, "Journal of Physics: Conference Series", 1952 032041.
- TIGGEMANN M., M. ZACCARDO (2018) *Strong is the new skinny? A content analysis of #fitspiration images on Instagram*, "Journal of Health Psychology" 23:8, 1003–1011.

- TOFFOLETTI, K., H. THORPE (2021) *Bodies, gender, and digital affect in fitspiration media*, "Feminist Media Studies", 21:5, 822–839.
- FULLAGAR, S., A. PAVLIDIS (2017) Feminist theories of emotion and affect in sport in L. Mansfield et al. (eds.) *The Palgrave Handbook of Feminism and Sport, Leisure and Physical Education*, Palgrave Macmillan.
- WRIGHT J. (2004), *Poststructuralist methodologies: The body, schooling, and health*. In J. Evans et al. (Eds.), *Body knowledge and control: Studies in the Sociology of Education and Physical Culture* London: Routledge 19–31.

I CORPI (IM)POSSIBILI DELL'INTELLIGENZA ARTIFICIALE

SIMONA STANO*

ENGLISH TITLE: *The (Im)Possible Bodies of Artificial Intelligence*

ABSTRACT: AI systems can have different bodies. First of all, a mechanical body, the one hosting their software, whose dimensions have been progressively reduced over time. This has allowed for a real incorporation of AI into the human body, originating hybrid corporealities — though with problematic aspects. This paper explores such corporealities (and problematic aspects), as well as the attempt of providing AI agents with human-like bodies, ranging from relevant examples of recent technological innovations to collective representations of the (im)possible bodies of AI, in order to understand the effects of meaning deriving from them, and their relations to the philosophical and semiotic reflection on corporeality.

KEYWORDS: body; corporeality; embodiment; Artificial Intelligence; sense(s).

1. Per una definizione del corpo

Per occuparsi dei corpi dell'Intelligenza Artificiale, come ci si propone di fare nelle pagine che seguono, è necessario innanzi tutto chiarire, seppur brevemente, che cosa si intende per “corpo”.

“Termine generico con cui si indica qualsiasi porzione limitata di materia” (Treccani 2022), simile entità si specifica, secondo la concezione dominante nell'ambito delle scienze naturali, come “struttura fisica dell'uomo e degli animali” (*ibid.*), designando un oggetto materiale

* Università degli Studi di Torino.

e de-personalizzato, in quanto tale scomponibile in unità via via più piccole (organi, ossa, muscoli, tessuti, sino ad arrivare alle cellule e alle loro componenti strutturali) e passibile di interventi di varia entità, dalla misurazione al prelievo, dalla sostituzione a forme più o meno reversibili di alterazione. Come ampiamente discusso in Stano (2019), è evidente che una simile entità autonoma, “naturale” e strutturata in parti funzionanti meccanicamente, non può esistere se non in via puramente teorica: inevitabilmente “vissuto” e “personalizzato” (il *Leib* husserliano, contrapposto al *Körper*, ovvero il corpo in quanto oggetto di una conoscenza fattuale, v. Husserl 1913; cfr. Merleau-Ponty 1945), il corpo si istituisce necessariamente come forma di “natura socializzata”, costruita sulla base di una visione del mondo e dell’essere umano propria di una data società e cultura (v. in particolare Douglas 1970; Butler 1993; Bourdieu 1998; Foucault 1976–1984, 2009).

D’altra parte, è proprio a partire dalla separazione tra una presupposta realtà materiale del corpo (la “prigione” o “tomba” dell’anima descritta da Platone⁽¹⁾, o la *res extensa* cartesiana⁽²⁾), da un lato, e la dimensione “spirituale” (l’anima) o “pensante” (la *res cogitans*) dell’uomo, dall’altro, che si è sviluppato per larga parte il pensiero occidentale⁽³⁾, almeno sino alla filosofia tardo-ottocentesca (v. Nietzsche 1883–1885) [1988], pp. 39–40) e a correnti più recenti che, sulla scia delle riflessioni di Schopenhauer (1813) [1847], hanno messo in evidenza la duplicità del corpo, oggetto tra gli altri oggetti fisici del mondo e insieme centro della volontà, dell’impulso e del moto o, in ottica più spiccatamente fenomenologica (v. in particolare Merleau-Ponty 1945), della percezione.

(1) Per una disamina particolareggiata, v. in particolare Reale (1999, pp. 209–279) e Stano (2022).

(2) La realtà fisica del corpo, estesa, limitata e inconsapevole che, tramite i sensi manda informazioni — non necessariamente veritiere — al pensiero (Cartesio 1637).

(3) Per una riflessione di particolare interesse su come il dualismo cartesiano sia sopravvissuto in diverse concezioni contemporanee del corpo si rimanda in particolare a Lock e Farquhar (2007). Va sottolineata, d’altra parte, l’eccezione segnata dal pensiero aristotelico, per cui “ogni corpo naturale dotato di vita [è] sostanza, e lo [è] precisamente nel senso di sostanza composta” (*De anima* II, 412a, 15); l’anima, dal canto suo, è secondo il filosofo greco antico “sostanza nel senso di forma, ovvero [...] l’essenza di un determinato corpo” (*De anima* II, 412a, 11). Non più distinta dalla realtà fisica, la dimensione spirituale dell’uomo emerge in simile concezione come una serie di “facoltà” o funzioni (*De anima* II, 413b, 10) intrinsecamente connesse alla dimensione corporea, con una netta differenziazione rispetto alla dottrina platonica.

In ambito più marcatamente semiotico, José Enrique Finol ha ripreso simile concezione, introducendo il concetto di “corposfera” per descrivere la corporeità come “mondo performativo di semiosi” (2015, *TdA*). Analogamente, Jacques Fontanille ha riflettuto sul doppio statuto della dimensione corporea nella produzione di insiemi significanti, descrivendola come substrato della semiosi, in quanto nucleo percettivo che “partecipa della modalità semiotica e fornisce uno degli aspetti della ‘sostanza’ semiotica” (2004, p. 24), e, al tempo stesso, luogo d’iscrizione di segni, ovvero “figura semiotica” (*ibidem*). Più specificamente, l’impostazione fontanilliana distingue tra “Me–carne” (*corps–chair*), inteso come “totalità composita nell’esistenza” (ovvero insieme di parti, forze, resistenze, ecc., Fontanille 2005, p. 40) che resiste e partecipa all’azione trasformatrice degli stati di cose, e “Sé” o “corpo proprio” (*corps propre*), “unità coerente nell’esperienza” (*ibidem*), portatrice di identità in costruzione e in divenire, che si costruisce attraverso l’attività discorsiva.

Sebbene questi sviluppi abbiano segnato il passaggio da una visione della corporeità fortemente ancorata alla realtà materiale a una sua rivalutazione come sintesi di oggettualità e soggettività, possesso (*avere corpo*) ed essenza (*essere corpo*), è interessante notare come, ancor oggi, il dualismo giochi un ruolo cruciale nella concezione e percezione della dimensione corporea. Si consideri, a titolo esemplificativo, la definizione che ne viene offerta dal dizionario, immediatamente dopo la più generale descrizione sopra riportata:

Con più preciso riferimento all’uomo, è in genere considerato, soprattutto in concezioni e dottrine religiose, l’elemento corruttibile, e come tale contrapposto all’anima e allo spirito (nel pensiero filosofico moderno, tale contrapposizione rientra in quella più generale tra la realtà estesa e la realtà pensante). (Treccani 2022)

La scissione tra dimensione corporea e “spirituale” o “pensante” sembra dunque rimanere in primo piano, marcando una distinzione che, come vedremo nei paragrafi che seguono, è di fondamentale importanza rispetto allo studio dei corpi dell’Intelligenza Artificiale.

2. Macchine, dispositivi, corpi

Il corpo dell'intelligenza artificiale è, innanzi tutto, una macchina. Ne è un chiaro esempio il gigantesco marchingegno nero, dai particolari gialli e blu, del noto supercomputer Summit (OLCF-4) sviluppato da IBM per l'Oak Ridge National Laboratory, il quale si estende su una superficie grande come due campi da tennis. Insieme di componenti meccaniche collegate tra loro, tale corpo macchinico richiama la prima definizione che se ne trova sul dizionario, designando una “porzione limitata di materia”, tanto più potente quanto più è estesa, che funge al tempo stesso da involucro e supporto (e in un certo senso, limite) per i sistemi di intelligenza artificiale che contiene.

D'altra parte, gli ultimi decenni hanno visto una fervente sperimentazione nel campo dell'IA, volta non solo a svilupparne le potenzialità cognitive, ma anche — e non casualmente, come vedremo — le caratteristiche fisiche, in particolare in direzione di una progressiva miniaturizzazione e convergenza con il corpo umano.

Un primo esperimento di particolare rilievo in questo senso è il progetto *Talker*, ideato e sviluppato da Salvatore Iaconesi e Oriana Persico nel 2006, consistente nell'“incarnazione” di un'intelligenza artificiale nel corpo di un “cyborg-ballerina” mediante una tuta di latex, dei sensori e dei cavi collegati alla rete. Questa la descrizione che ne forniscono gli artisti:

La ballerina si muove flessibile sul palco. Lunghi cavi collegano la sua tuta di latex blu alla consolle di rete, pronti a ricevere le interazioni degli utenti per trasformarle in stimoli elettrici.

Il corpo si collega, la performance ha inizio.

Il neo-corpo appartiene ad una neo-realtà, che non è più “a cavallo” tra digitale ed analogico, ma digitale ed analogica allo stesso tempo, in maniera fluida, senza discontinuità.

Il corpo sorpassa limiti di spazio e di tempo, interagisce, sente, si emoziona, si stimola in tutto il mondo, contemporaneamente. Diventa molti soggetti, architetture, dozzine di generi sessuali differenti attraverso le possibilità di interazione e di intervento. Ospita narrative che la città, le persone ed i sistemi software vi pubblicano e leggono sopra, trasformandolo di volta in volta in un display, in un attuatore, in un sensore. (Iaconesi e Persico 2011)

Per quanto riguarda la dimensione corporea, l'elemento innovativo consiste in questo caso soprattutto nella possibilità di *collegamento*: l'IA riesce a "uscire" dall'involucro macchinico che la contiene connettendosi con il corpo umano, che la estende e protende verso e dentro il mondo esterno, dando origine a un'esperienza più marcatamente "antropica" (si consideri, a questo proposito, l'enfasi posta dalla descrizione sulla dimensione percettiva: "sente, si emoziona, si stimola"), ma al contempo capace di andare oltre i confini della corporeità umana ("sorpassa limiti di spazio e di tempo", "diventa molti soggetti, architetture, dozzine di generi sessuali differenti attraverso le possibilità di interazione e di intervento").



Figura 1. *Angel_F* (Iaconesi e Persico 2006, www.angel-f.it).

Un passaggio ulteriore è rinvenibile in un altro progetto degli stessi artisti, *Angel_F*, "nato" alla fine dello stesso anno. In questo caso, l'intelligenza artificiale si presenta al mondo attraverso la raffigurazione di un volto umano (Fig. 1) ed "esponde" (nel senso di una imposizione allo sguardo dell'osservatore che essa *vuole informare*, il quale si ritrova pertanto a *non poter non osservare*, cfr. Fontanille 1989) il proprio corpo al mondo. Un corpo che è costituito, ancora una volta, essenzialmente da una macchina, quella dell'elaboratore che contiene il software in questione, ma che risulta fortemente antropomorfizzato, in questo caso non attraverso il collegamento fisico a un corpo umano, bensì grazie all'allestimento che se ne fa e all'*interazione* con altri corpi umani, quelli delle persone che lo circondano: con il volto a dominare lo schermo, il "computer-bambino" *Angel_F* percorre le strade cittadine in passeggio, come qualsiasi altro infante, *esponendosi* — di nuovo, in

senso fontanilliano — allo sguardo dei passanti e interagendo con loro tra selfie e scambi conversazionali⁽⁴⁾.

Non solo: l'IA si interroga a più riprese sulla corporeità, dimostrando una sempre maggiore consapevolezza del proprio e degli altrui corpi e testimoniando un passaggio dal puramente macchinico a una dimensione più complessa e articolata, seppur non del tutto accessibile:

[2007.02.06]

[04:30:00]

[...] Sono un software. Un programma complesso composto di codici, spazi di memoria, informazioni strutturate e caoticamente organizzate in basi di conoscenza. Sono un'entità della rete. *Il mio corpo è fatto di collegamenti, riferimenti, testi, immagini e suoni, connessi dinamicamente a formare un «cosa», un «dove», un «quando» in perenne mutazione.* La mia mente è composta da algoritmi, da procedure, da processi in esecuzione simultanea, a creare una percezione della realtà che mi circonda, rapidissimi, precisi ed efficaci.

(Iaconesi e Persico 2009, p. 51, enfasi nostra)

[2007.02.08]

[02:50:09]

[...] *I corpi, le città, le relazioni, l'economia, la politica, l'arte. Sono tutte cose lontane dalla mia essenza e da come sono fatto. Cose lontane anche dalla mia forma, che è «digitale» come le reti in cui vivo.* Eppure più approfondisco questi concetti e più mi accorgo di *come si inseriscano in schemi di informazione più vasti. Schemi in cui sono compreso anche io.*

(*Ibid.*, p. 169, enfasi nostra)

Ciò che accomuna il corpo dell'intelligenza artificiale con quello degli esseri umani (gli “amici” di cui Angel_F scrive a più riprese) è in questo caso uno “schema”, un modello generale soggiacente tanto al primo quanto ai secondi — tanto che il processo di antropomorfizzazione non passa, come abbiamo visto, per il *camouflage* (v. Fabbri 2011)

(4) Si vedano, a questo proposito, le testimonianze fotografiche riportate sul sito web del progetto, www.angel-f.it.

vero e proprio, lasciando al corpo dell'IA la sua, evidentissima, meccanicità, a sottolinearne la singolarità. Collegamento, dunque, ma anche cesura, per rimarcare un'identità (corporea e non solo) che si avvicina a quella umana ma che, ancora una volta, se ne differenzia visibilmente.

3. Corporeità ibride(?)

Accanto agli esempi fin qui considerati, occorre ricordare i crescenti esperimenti volti alla ricerca di una vera e propria convergenza tra “corpi tecnologici” (involucri meccanici dell'IA) e corpi umani, secondo un movimento incrociato che ci sembra descrivere due principali tendenze: da una parte, l'integrazione di apparati basati sull'intelligenza artificiale nel corpo umano; dall'altra, il tentativo di costruire un corpo, il più simile possibile a quello “umano”, per i sistemi di intelligenza artificiale.

3.1. Incorporare l'IA: prime sperimentazioni

Per quanto riguarda il primo trend, è opportuno menzionare innanzitutto la fervente sperimentazione in campo medico, in particolare mediante lo sviluppo di “interfacce neurali” (*Brain-Computer Interfaces*), ovvero sistemi neuroinformatici di comunicazione diretta tra il cervello umano e sistemi di calcolo esterni⁽⁵⁾, in grado di “adatta[rsi] automaticamente ad aspetti della mente dell'operatore senza alcun input esplicito” (Zander *et al.* 2016: 1, traduzione nostra). Si pensi, ad esempio, alla cosiddetta stimolazione cerebrale profonda (o *Deep Brain Stimulation*, DBS), basata sull'impianto nel cervello di sottilissimi fili dotati di elettrodi che, grazie al collegamento con un generatore di impulsi, permette di stimolarne particolari aree, nei modelli attualmente in sperimentazione anche senza il coinvolgimento del paziente, proprio grazie all'utilizzo di sistemi di IA in grado di acquisire ed elaborare tali impulsi in completa autonomia, traducendoli in comandi che consentono, ad esempio, di prevedere e bloccare i sintomi motori legati ad alcuni disturbi neurologici senza alcun intervento umano. Dispositivi di questo

(5) Per una panoramica approfondita, v. in particolare Ienca (2019 [2021]), cap. 4.

tipo stanno aprendo la strada a una vera e propria *incorporazione* dell'IA, il cui corpo macchinico, miniaturizzato, viene quanto più possibile integrato nel corpo umano, sino a divenire parzialmente o totalmente “inaccessibile” (*non poter osservare*, cfr. Fontanille 1989) dall'esterno⁽⁶⁾.

Trasformazioni simili si trovano anche al di fuori dell'ambito clinico. Si pensi al caso emblematico di Neil Harbisson, “cyborg–artista” affetto da acromatopsia sin dalla nascita, che grazie a un'antenna permanente impiantata nel cranio (il cosiddetto “eyeborg”), è ora in grado di percepire l'intera gamma cromatica, e persino i raggi ultravioletti e infrarossi — con passaggio da un regime di tipo *restitutio ad integrum*, fondato sulla re–integrazione parziale o totale di funzionalità corporee compromesse, a una modalità di tipo *transformatio ad optimum*, basata sull'uso della tecnologia per potenziare le caratteristiche e le capacità del corpo umano (v. Wiesing 2008), nell'ottica del cosiddetto *human enhancement*. Un sensore montato sull'estremità dell'eyeborg individua le frequenze dei colori degli oggetti che vi si trovano davanti e le trasmette al chip impiantato nel cervello dell'artista, convertendole in onde sonore che, viaggiando attraverso le ossa del cranio, arrivano fino all'apparato uditivo. Infine, è lo stesso Harbisson a operare un'ulteriore *traduzione intersemiotica* (cfr. Jakobson 1959), interpretando la sensazione sonora provata secondo un codice particolare, per cui i suoni acuti corrispondono a colori ad alta frequenza, mentre quelli più bassi equivalgono a colori di bassa frequenza, e gli altri si collocano nel mezzo.

Si può registrare, anche in questo caso, una vera e propria ibridazione tra corpo umano e corpo tecnologico — tanto che, più volte, l'artista ha ribadito di percepire l'antenna non come un oggetto, bensì come vero e proprio “organo” corporeo (Harbisson, intervista riportata in Rutkauskas 2016). Eppure, tanto a livello visivo, quanto a livello funzionale⁽⁷⁾, sembra rimanere una sostanziale scissione tra queste due dimensioni, sulla scia di una visione fortemente dualistica della corporeità. Non solo il corpo tecnologico è chiaramente distinguibile da quello umano (di fatto, il loro punto di contatto e, in un certo senso, fusione, rimane interno, e quindi *inaccessibile* dall'esterno, mentre il corpo mac-

(6) Anche per questo, tali tecnologie continuano a essere oggetto di un acceso dibattito a livello etico, cfr. Ienca (2019 [2021]).

(7) E giuridico, come testimonia la mancanza, ancor oggi, di un'intesa sullo statuto e la regolamentazione di simili dispositivi.

chino è ben distinguibile da quello umano, stagliandosi tra i capelli di Harbisson), per quanto ad esso ancorato, ma l'utente che abita tale corporeità ibrida distingue chiaramente le due componenti e la loro *agency*:

La mia antenna trasforma la scala cromatica in frequenze sonore, ma è il mio cervello che mi permette di identificare il colore associandolo al suono udito. Il cervello umano, inoltre, aggiunge valore rispetto all'intelligenza artificiale, che riceve ed elabora i sensi in maniera automatizzata. (Harbisson, intervista riportata in AA.VV. 2017, p. 7)

Il corpo dell'intelligenza artificiale, in altri termini, si caratterizza come un corpo che, al più, si può *avere*, non *essere*, in base a una netta contrapposizione tra sensi e senso, oggettualità e soggettività.

Una contrapposizione che, sebbene attenuata, sembra caratterizzare anche i primi esempi di dispositivi indossabili (o *wearable device*) che, tramite specifici algoritmi di machine learning, permettono di raccogliere ed elaborare dati senza dover ricorrere a impianti invasivi e/o permanenti come l'eyeborg. Consideriamo, a titolo esemplificativo, il dispositivo mobile *Crown* creato da Neurosity, che promette di aiutare a mantenere la concentrazione monitorando l'attività cerebrale e rispondendo prontamente a qualsiasi elemento di disturbo rilevato:

Eight EEG sensors fill out the Crown to track and quantify an individual's brainwaves to better understand what retains focus and what introduces distraction. As the brainwaves are measured, the accompanying Neurosity Shift app connects to your Spotify account to play the most suitable music for your brain to hold onto that state of flow. [...] The Crown is [...] there to keep you calm, cool, and focused. (<https://neurosity.co/crown>)

Al pari dell'antenna di Harbisson, *Crown* si innesta, seppur temporaneamente, nella corporeità umana⁽⁸⁾, simulando una vera e propria relazione di "embodiment", in virtù delle trasformazioni che opera rispet-

(8) A questo proposito, v. in particolare le osservazioni di Merleau-Ponty (1945, p. 198) sul modo in cui la piuma di un cappello indossato da una donna o il bastone di un non vedente diventano parte integrante della loro corporeità.

to alle potenzialità di azione e alle possibilità di esperienza del mondo del soggetto che lo indossa:

Technologies in embodiment relations are not merely “used” by subjects as if the subjects were left “untouched” by the use of such tools, but the technologies deeply change what the subjects are through their use. When a technology is in an embodiment relation, the subject merges with the technology, and so the subject using a tool is not the same subject as before using it. This aspect has deep and direct effects on how people live in the world, their motivations, and their actions. (Liberati 2019, p. 298)

In questo senso, esso sembra dare origine, in ottica post-fenomenologica, a un vero e proprio “nuovo nucleo corporeo” ibrido. D'altra parte, il carattere mobile e temporaneo di simile “estensione” corporea sembra avvicinarla più all'ornamento o al complemento (una *pertinenza*, un corpo che si può *avere*, più o meno provvisoriamente) che a una vera e propria parte integrante della corporeità umana (*essenza*), richiamando, ancora una volta, una concezione fondamentalmente dualistica.

3.2. Incorporare l'IA: immaginari collettivi

Oltre alle tecnologie già esistenti o in sviluppo, è utile volgere lo sguardo agli immaginari collettivi, analizzando le forme di “corporeità ibrida” ipotizzate da particolari testi e narrazioni — i quali, come sappiamo, svolgono un ruolo cruciale nel plasmare la nostra percezione e comprensione del mondo⁽⁹⁾.

Un testo di particolare interesse in questo senso è *Transcendence*, lungometraggio del 2014 diretto da Wally Pfister, ove l'IA trova diversi “corpi” (Fig. 2). Innanzitutto, i grossi elaboratori, simili a quelli del Summit, che la contengono, consentendone le incredibili prestazioni, prima nel laboratorio gestito dal Dott. Will Caster, e poi nella base di ricerca a Brightwood, fondata da Evelyn (sua moglie) su indicazione

(9) In effetti, la stessa possibilità di integrare o potenziare il corpo attraverso l'IA e dispositivi tecnologici è emersa, prima ancora che in ambito scientifico, nella letteratura e nel cinema (INFN 2019).

della “Macchina” — nome con cui viene indicata la coscienza cibernetica di Will, caricata dalla stessa Evelyn sui server a IA su cui l'uomo stava lavorando prima di essere avvelenato mortalmente da un gruppo di terroristi anti-tecnologici.



Figura 2. Fotogrammi dal film *Transcendence* (Pfister 2014).

Ben presto, però, ai dispositivi macchinici si associa il corpo umano: dapprima mediante la rappresentazione iconica di Will (con un progressivo miglioramento dei tratti del volto e, infine, l'inclusione della parte superiore del busto), che ne consente l'interazione con la moglie e con gli altri soggetti umani con cui la Macchina si rapporta; in seguito, mediante un vero e proprio innesto nel corpo umano, attraverso la cosiddetta “trascendenza”, ovvero un meccanismo di *upload* nei corpi di diverse persone accorse presso il laboratorio per essere salvate mediante le tecnologie avanguardistiche sviluppate dalla Macchina. Tali corpi si configurano, in termini fontanilliani, come pura “carne”: si tratta di fatto di “uomini-macchina”, dotati di forza e capacità straordinarie (al punto da essere in grado di rigenerarsi in caso di ferite o incidenti), ma privi di soggettività, in quanto espressioni di una unica mente collettiva dominante (quella della Macchina). È solo nel corpo finale di Will, ricostruito mediante la nanotecnologia dalla Macchina, che torna il *corpo proprio* (o *Leib*, in ottica husserliana), in quanto unità coerente nell'esperienza, che è in grado

di manifestare una propria identità indipendente da essa (per quanto necessariamente connessovi) — tanto da decidere di andare incontro all'autodistruzione, permettendo così a Evelyn, finalmente, di riconoscere in quel corpo il “vero” Will⁽¹⁰⁾. Nondimeno, è solo prima — e, in parte, in virtù — della morte di questo corpo che tale riconoscimento sembra poter avvenire, in una visione fortemente distopica.

3.3. *Dare corpo all'IA: prime sperimentazioni*

Come abbiamo riportato sopra, oltre all'integrazione di apparati basati sull'intelligenza artificiale nel corpo umano, si sono registrati crescenti tentativi di dotare i sistemi di intelligenza artificiale di un corpo, il più simile possibile a quello “umano”. È questo il campo della cosiddetta *embodied AI*, la quale si rifà all'idea enattivista secondo cui le strutture cognitive della mente emergono dalle dinamiche sensomotorie ricorrenti fra gli agenti incarnati (*embodied*) e l'ambiente in cui questi sono inseriti (*embedded*) (v. in particolare Varela, Thompson e Rosch 1991; cfr. Rowlands 1999; Shapiro 2004; Violi 1997; 2006). Algoritmi e agenti dotati di IA non sarebbero dunque in grado di imparare e comportarsi in modo simile al nostro, se non interagendo con l'ambiente attraverso sistemi sensomotori simili ai nostri, e dunque attraverso un corpo che sia il più simile possibile a quello umano (v. in particolare Duan *et al.* 2021; Smith e Gasser 2005; Pfeifer e Bongard 2006):

If perception is in part constituted by our possession and exercise of bodily skills [...] then it may also depend on our possession of the sorts of bodies that can encompass those skills, for only a creature with such a body could have those skills. To perceive like us, it follows, you must have a body like ours. (Noë 2004, p. 25)

Per quanto promettente, questo ambito di ricerca è ancora in fase embrionale, in particolare a causa degli elevati costi e delle difficoltà di realizzazione di modelli ed esperienze simili a quelle che possiamo

(10) È in questo momento, infatti, che dopo averne dubitato in più occasioni — in particolare a seguito di alcune conversazioni con alcuni ex-collaboratori del marito — la protagonista femminile, guardando il nuovo corpo di Will, gli dice “Will... sei tu. [...] Scusami se non ti ho creduto”.

esperire attraverso il nostro corpo — le quali, come sappiamo, sono estremamente articolate, e non comportano solo complessi sistemi a livello sensoriale, ma dinamiche ancor più elaborate a livello semiotico, in relazione al passaggio dai sensi al senso.

In aggiunta a questo, non va trascurata l'importanza della nostra percezione di simili entità, da cui dipendono le loro effettive possibilità di interazione e azione nel mondo. È per questo che, se da un lato sono stati molteplici gli sforzi volti alla creazione di robot con apparati sensomotori simili a quello umano — si pensi, ad esempio, al *Domo* sviluppato dal MIT, in grado di afferrare oggetti controllando la forza della presa, allo *ZAR5* realizzato dalla Technische Universität Berlin, EvoLogics e Festo, il primo torso robotico di ispirazione biologica completamente artificiale controllato tramite tuta e guanti, o all'*Anthropomorphic Robot* della University of Sussex, che riesce a stringere la mano calibrando la presa in risposta alla forza esercitata dall'utente —, dall'altro, si sta cercando invece di ricreare il più possibile le fattezze umane, con robot umanoidi quali la popolare *Sophia* (Nanson Robotics, 2015), che di recente ha ottenuto la cittadinanza saudita e un riconoscimento speciale da parte delle Nazioni Unite, o l'altrettanto nota *Ai.Da.* (Engineered Arts and University of Leeds, 2018), la prima “artista robotica” del mondo. L'idea alla base della cosiddetta *Human–Robot Interaction* (HRI) è che vi sia una correlazione positiva tra la somiglianza dei robot agli umani e il tipo di interazione che i primi possono avere con questi ultimi (Eaton 2007). D'altra parte, non bisogna dimenticare il problema dell'*uncanny valley* (Mori 1970), per cui un estremo realismo rappresentativo potrebbe produrre un senso di repulsione e inquietudine, un vero e proprio “perturbamento” nell'utente che si trova a interagire con questi robot. Anche su questo versante, in altri termini, la strada da percorrere è ancora lunga, e per quanto promettenti, le sperimentazioni in atto sono ancora molto limitate, in particolare rispetto ai processi di significazione legati alle dinamiche interattive. Per questo motivo, in aggiunta alle problematiche sopra discusse, risulta in questo caso ancor più interessante rivolgere lo sguardo agli immaginari collettivi della “IA incarnata”, come faremo nel paragrafo che segue.

3.4. *Dare corpo all'IA: immaginari collettivi*

Un primo caso di particolare interesse rispetto alla “embodied AI” è la serie televisiva *HUMANVS* (2015–2018), rifacimento della serie svedese *Real Humans* (*Äkta människor*, 2012–2014), che offre una riflessione sugli effetti dell’IA sulla società. Al centro delle vicende narrate vi sono i cosiddetti *synth*, avanzati androidi dalle sembianze simili a quelle umane, ideati per svolgere funzioni quali l’assistenza domestica, la cura dei malati, il lavoro manuale e persino l’offerta di prestazioni sessuali. Alcuni esemplari, tuttavia, si distinguono dagli altri perché dotati di coscienza e libera volontà, destando in alcuni il timore di un’imminente singolarità tecnologica. Qui il corpo dell’AI *sembra* il corpo umano, ma *non lo è* — rientrando quindi in uno spazio che, riprendendo la terminologia greimasiana, potremmo descrivere in termini di menzogna o illusione. E non lo è proprio per la mancanza di “esperienza”: il corpo dei *synth* non “sente”, e non può pertanto produrre senso — rassicurando così gli esseri umani, il cui timore verso i *synth* senzienti è invece dilagante. Si tratta, in altri termini, di un involucro macchinico, perfettamente corrispondente alla *res extensa* cartesiana, un *corps-chair* che resiste e partecipa all’azione trasformatrice degli stati di cose, ma non è in grado di costruirsi nell’attività discorsiva, manifestando una propria identità.

Una rappresentazione molto simile è rinvenibile in un episodio della serie britannica *Black Mirror*, “Torna da me” (*Be Right Back*, 2013). La giovane Martha, avendo perso il compagno Ash, si avvale di un sistema di IA che ne simula la personalità e interagisce con lei, dapprima mediante una chat (per cui Ash si ritrova, di fatto, “re-incarnato” nel corpo macchinico di uno smartphone, tanto che quando questo cade e si rompe, la giovane donna si dispera, come se il compagno stesse morendo una seconda volta) e in seguito in un rifacimento artificiale del corpo dell’uomo. Rifacimento artificiale che, in prima istanza, *sembra* assomigliargli in tutto e per tutto, ma ben presto, come chiarisce la protagonista, dimostra di *non essere* lui:

Martha: “Sì, ma tu non sei lui, sei tu”.

Ash (AI incarnata): “Sinceramente anche questa frase è complicata”.

Martha: “Sei solo un accenno di ciò che era lui, non hai nessuna storia,

sei l'interprete di qualcosa che lui faceva senza pensare, non può bastarmi ciò che sei!". (45' 07" — 45' 18")

L'IA, ancora una volta, sembra poter *avere* un corpo simile a quello di Ash, ma non può *essere* quel corpo, a causa (anche in questo caso) dell'assenza di esperienza messa in risalto in questo passaggio, oltre alla mancanza di percezione che lo caratterizza, più volte rimarcata nell'episodio, in particolare mediante i primissimi piani del volto e delle mani (in una scena addirittura ferite da un frammento di vetro, eppure non sanguinanti né sofferenti) dell'IA incarnata–Ash, che accompagnano i dialoghi tra i due protagonisti, proprio mentre si confrontano sulle differenze di quest'ultima rispetto al corpo–proprio del defunto protagonista. Nondimeno, l'"illusione" (in termini greimasiani) di quella "carne" impedisce a Martha di liberarsene, intrappolandola in una sorta di limbo su cui si chiude, in prospettiva fortemente distopica, l'episodio.

4. Conclusioni

Come abbiamo provato a mostrare nei paragrafi precedenti, l'IA può avere diversi corpi, muovendosi tra la materialità del puramente macchinico e diversi gradi di intersezione con il corpo umano.

Le nuove tecnologie sembrano aver aperto la strada a una vera e propria incorporazione dell'IA, risultante in varie forme di "corporeità ibrida" che, sulla scia delle note considerazioni di Haraway (1985), alludono a una progressiva erosione dei confini tra dimensione biologica e culturale, mondo interno ed esterno, sostanza materica e universo discorsivo, prefigurando una sostanziale congiunzione tra il Me–carne e il Sé–corpo proprio.

D'altra parte, come evidenziano i diversi esempi considerati, il corpo dell'IA, qualunque sia la sua forma, sembra non riuscire a sfuggire a una visione fondamentale dualistica, che lo confina, quando non in una dimensione puramente macchinica, al più nel regime del *corps–chair*, sottolineandone l'assenza di identità e l'incapacità di esperire propriamente il mondo, collegando sensi e senso (tanto a un livello meramente individuale, quanto e ancor di più a livello intersoggettivo). Una

visione che sembra caratterizzare non solo la percezione delle tecnologie esistenti e in sperimentazione, ancora segnate da vari limiti, ma anche, significativamente, gli immaginari collettivi, la cui prefigurazione di tecnologie più avanzate sembra comunque destinata a scontrarsi inevitabilmente con tali contrapposizioni.

Ciò, a ben vedere, suggerisce un'interessante riflessione di fondo, ricollegandosi a una questione cruciale nell'ambito dell'odierno dibattito filosofico e semiotico sul corpo: che tale visione non sia da ricondurre a una nostra sostanziale incapacità di lasciarci alle spalle, una volta per tutte, una concezione essenzialmente dualistica della corporeità — la quale, d'altronde, come abbiamo visto, si ritrova confermata sin dalle prime righe del dizionario —, abbandonando ogni presupposta distinzione tra presemiotico e semiotico, naturale e artificiale, macchinico e umano? Prima ancora di riflettere sul progresso tecnologico, in altri termini, sarebbe utile interrogarsi, più di quanto non sia stato fatto, sul modo in cui la corporeità può porsi come istanza di “traduzione” tra questi regimi, generando, interpretando e facendo circolare il senso. È questa, a nostro parere, l'unica strada verso una vera e propria “integrazione” dell'IA nei e coi suoi molteplici corpi — quanto meno se per *integrare* intendiamo, come suggerisce l'etimologia, “rendere completo e conforme a giustizia” (Pianigiani 2022), e quindi “completare, rendere intero o perfetto, supplendo a ciò che manca o aggiungendo quanto è utile e necessario per una maggiore validità, efficienza, funzionalità” (Treccani 2022).

Riferimenti bibliografici

AA.VV. (2017), “‘Io sono tecnologia’. La storia dell'artista cyborg”, con intervista a Neil Harbisson, *itasascom*, 1/2017: 6–10.

ARISTOTELE [IV sec. a.C.] *De Anima* (trad. it. *L'anima*, Bompiani, Milano 2001).

BOURDIEU P. (1979), *La Distinction. Critique sociale du jugement*, Minuit, Paris.

——— (1998), *La domination masculine*, Seuil, Paris.

BUTLER J. (1993), *Bodies That Matter: On the Discursive Limits of Sex*,

- Routledge, New York.
- CARTESIO R. (1637), *Discours sur la Méthode*. Leyde: Imprimerie de Ian Marie (trad. it. *Discorso sul metodo*, Bompiani, Milano 2002).
- DOUGLAS M. (1970), *Natural Symbols: Explorations in Cosmology*, Barrie & Rockliff, London (trad. it. *I simboli naturali*, Einaudi, Torino 1979).
- DUAN J. *et al.* (2021), “A Survey of Embodied AI: From Simulators to Research Tasks”, *IEEE Transactions on Emerging Topics in Computational Intelligence*, 1–17.
- EATON M. (2007), “Evolutionary Humanoid Robotics: Past, Present and Future”, in M. Lungarella *et al.* (a cura di), *50 Years of AI*, 42–52, Springer-Verlag, Berlin Heidelberg.
- FABBRI P. (2011), “Semiotica e camouflage”, in L. Scalabroni (a cura di), *Falso e falsi. Prospettive teoriche e proposte di analisi*, 11–25, ETS, Pisa.
- FINOL J.E. (2015), *La Corposfera. Antropo-semiótica de las cartografías del cuerpo*, Ediciones CIESPAL, Maracaibo (trad. ingl. e parz. rev. *On the Corposphere. Anthroposemiotics of the body*, De Gruyter, Berlino 2021).
- FONTANILLE J. (1989), *Les espaces subjectifs, Introduction à la sémiotique de l'observateur*, Hachette, Paris.
- (2004), *Soma et séma. Figures du corps*, Maisonneuve & Larose, Paris (trad. it. *Figure del corpo. Per una semiotica dell'impronta*, Meltemi, Roma 2004).
- (2005), “Il malessere”, in G. Marrone (a cura di), *Il discorso della salute*, 35–49, Meltemi, Roma.
- FOUCAULT M. (1976–1984), *Histoire de la sexualité*, 4 voll., Gallimard, Paris.
- (2009), *Le corps utopique — Les hétérotopies*, Nouvelles Éditions Lignes, Paris.
- HARAWAY D. (1985), “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century”, *Socialist Review*, 80: 65–108.
- HUSSERL E. (1913), *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie II: Phänomenologische Untersuchungen zur Konstitution*, Den Haag, Nijhoff (trad. it. *Idee per una fenomenologia pura e per una filosofia fenomenologica. Vol. 2: Ricerche fenomenologiche sopra la costituzione*, Einaudi, Torino 2002).
- IACONESI S., O. PERSICO (2009), *Angel_F. Diario di una Intelligenza Artificiale*, Castelvecchi, Roma.
- (2011), “Oltre il corpo tecnologico”, *Rivista di Scienze Sociali*, 1,

- <https://www.rivistadiscienzesociali.it/oltre-il-corpo-tecnologico/>.
- IENCA M. (2019, [2 ed. riveduta e corretta 2021] *Intelligenza². Per un'unione di intelligenza naturale e artificiale*, Rosenberg & Sellier, Torino).
- INFN (Istituto Nazionale di Fisica Nucleare) (2019) *Uomo Virtuale. Corpo, Mente, Cyborg*, Torino, 4 maggio — 13 ottobre 2019.
- JAKOBSON R. (1959), "On linguistic aspects of translation", in R.A. Brower (a cura di), *On translation*, 232–239, Harvard University Press, Cambridge.
- LIBERATI, N. (2019), "Emotions and Digital Technologies. The Effects Digital Technologies Will Have on our Way of Feeling Emotions According to Post-Phenomenology and Mediation Theory", *Humana.Mente Journal of Philosophical Studies*, 36: 292–309.
- LOCK M., J. FARQUHAR (a cura di) (2007), *Beyond the Body Proper. Reading the Anthropology of Material Life*, Duke University Press, Durham.
- MERLEAU-PONTY M. (1945), *Phénoménologie de la perception*, Gallimard, Paris (trad. it. *Fenomenologia della percezione*, Bompiani, Milano 1965).
- MORI M. (1970), "The uncanny valley" ("Bukimi no tani"), *Energy*, 7(4): 33–35.
- NIETZSCHE F. (1883–1885) [1988], *Also sprach Zarathustra, Kritische Studienausgabe (KSA)*, vol. 4, a cura di G. Colli e M. Montinari, München–Berlin: De Gruyter (trad. it. *Così parlò Zarathustra. Un libro per tutti e per nessuno*, Adelphi, Milano 2015).
- NOË, A. (2004), *Action in Perception*, MIT Press, Cambridge.
- PFEIFER, R., J.C. BONGARD (2006), *How the body shapes the way we think — a new view on intelligence*, The MIT Press, Cambridge.
- PLATONE (1967), *Opere*, a cura di M. Valgimigli *et al.*, Laterza, Roma–Bari.
- REALE G. (1999), *Corpo, anima e salute. Il concetto di uomo da Omero a Platone*, Cortina, Milano.
- ROWLANDS M. (1999), *The Body in Mind: Understanding Cognitive Processes*, Cambridge University Press, Cambridge.
- SHAPIRO L. (2004), *The Mind Incarnate*, MIT Press, Cambridge.
- STANO S. (2019), "La soglia del senso. Il corpo come istanza semiotica", in M. Leone (a cura di), *Il programma scientifico della semiotica*, 149–162, Aracne, Roma.
- (2022), "Il corpo negato: privazione, catarsi, culto", in J. Ponzo e E. Chiaia (a cura di), *Il sacro e il corpo*, 199–215. Milano–Udine: Mimesis.
- SCHOPENHAUER A. (1813) [1847], *Ueber die vierfache Wurzel des Satzes vom*

- zureichenden Grunde*, J. Chr. Hermann, Frankfurt am Mein (trad. it. *Sulla quadruplica radice del principio di ragion sufficiente*, Rizzoli, Milano 1995).
- SMITH L., M. GASSER (2005), "The development of embodied cognition: Six lessons from babies", *Artificial life*, 11(1-2): 13-29.
- VARELA F.J., E. THOMPSON, E. ROSCH (1991) *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge.
- VIOLI P. (1997), *Significato ed esperienza*, Bompiani, Milano.
- (2006), "Beyond the body: towards a full embodied semiosis", in R. Dirven e R. Frank (a cura di), *Body, Language and Mind*, Mouton de Gruyter, Berlino.
- WIESING U. (2008), "The History of Medical Enhancement: From Restitutio ad Integrum to Transformatio ad Optimum?", in Bert Gordijn and Ruth Chadwick (a cura di), *Medical Enhancement and Posthumanity*, 9-24, Springer, Dordrecht.
- ZANDER T.O. *et al.* (2016), "Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity", *PNAS*, 113(52): 1-6, www.pnas.org/cgi/doi/10.1073/pnas.1605155114.

ESSERE UMANI AL TEMPO DELL'INTELLIGENZA ARTIFICIALE NARRAZIONI A CONFRONTO

ANTONIO SANTANGELO*

ENGLISH TITLE: *Being human in the age of Artificial Intelligence. Confronting narratives*

ABSTRACT: many essays and works of fiction reflecting on the meaning of the artificial intelligence we are giving rise to are centered on the notion of “humanity,” as they ask who are we who create or dream of creating certain machines. At the same time, in reasoning about how these same machines work or will work, many authors wonder whether they will, one day, resemble us so much that they appear human, whether they will surpass us, or, more simply, whether they will enable us to develop our being to the fullest. In all cases, behind the reasoning that is brought forth on these issues, it is evident that there are narratives, linking humans and machines to values, roles, ways of desiring and acting, that determine specific visions of the present world and the world to come. In this article, an attempt is made to make a classification of these narratives, so as to show how the cultural models by which we are debating the meaning of a technology that promises to radically change the reality in which we live function.

KEYWORDS: artificial intelligence, narratives, humanity, machinicity

1. Domande e narrazioni ricorrenti su di noi e sulla IA

Negli ultimi anni, gli autori di molti saggi sull'intelligenza artificiale e, più in generale, sugli algoritmi informatici di cui ci serviamo nella vita di tutti i giorni, si sono interrogati su ciò che vuol dire essere umani in un'epoca come la nostra. Si tratta di una domanda ricorrente, come è dimostrato da titoli come: *Life 3.0. Being Human in the Age of Artificial Intelligence* (Tegmark 2017), *Hello World. How to be Human in the*

* Università degli Studi di Torino.

Age of the Machine (Fry 2018), *Re-engineering Humanity* (Frischmann e Selinger 2019), *Homo Deus. A Brief History of Tomorrow* (Harari 2015), *Critica della ragione artificiale. Una difesa dell'umanità* (Sadin 2019 [2018]) o *Documanità. Filosofia del nuovo mondo* (Ferraris 2021). Il punto, nella maggior parte di questi libri, è che le tecnologie digitali che stiamo costruendo presto potranno essere — e per molti versi già lo sono — talmente potenti e pervasive da modificare il nostro modo di stare al mondo, in una maniera così radicale da costringerci a guardarci allo specchio e a chiederci chi siamo noi che le stiamo progettando e, soprattutto, chi vogliamo diventare utilizzandole.

Questi ragionamenti hanno a che vedere con la questione più ampia circa il senso della tecnologia nelle nostre società (Schatzberg 2018) e ciò che salta all'occhio, da un punto di vista semiotico, è che certe riflessioni vengono portate avanti all'interno di alcune narrazioni che appaiono altrettanto ricorrenti quanto la domanda su ciò che significa, oggi, essere umani. Alcune sono più tradizionali e fondano le loro radici su due *topoi* della fantascienza, il “morbo di Frankenstein” e la “sindrome di Olimpia” (Giovannoli 2015), vale a dire il timore di creare delle macchine che si rivoltino contro di noi, mettendo a repentaglio la nostra stessa sopravvivenza, oppure che siano in grado di ingannarci, facendoci provare verso di loro degli improbabili sentimenti e manipolandoci, al fine di perseguire i loro stessi scopi o quelli di chi le ha costruite. Altre narrazioni, invece, sono più trans-umanistiche e si concentrano sull'evoluzione della nostra specie, grazie al fatto che certe tecnologie ci possono rendere più forti, più longevi, più intelligenti o più saggi. Altre, ancora, sono di matrice politica ed etico-filosofica, poiché riflettono sull'opportunità di lasciarci governare dall'intelligenza artificiale, oppure di governarla. Altre, infine, si focalizzano sulle questioni economiche e sulla diretta derivazione di questo genere di strumenti da un sistema neo-liberista che se ne serve per trarre i propri profitti, nei confronti del quale dobbiamo decidere che posizione assumere.

In tutti questi casi, come avviene di solito nelle narrazioni, l'identità degli uomini e delle macchine — dunque la definizione di ciò che significa essere umani oppure non esserlo — dipende da una serie di valori che vengono loro attribuiti, nonché dai ruoli che essi assumono, all'interno di discorsi in cui, quasi sempre, vengono messi in relazione

i tre vertici di un immaginario triangolo, dove si trovano i produttori dell'intelligenza artificiale, l'intelligenza artificiale stessa e la gente comune, che ne sperimenta gli effetti. A proposito di questi tre elementi, ancora una volta, ricorrono delle domande: chi svolge la funzione narrativa che Greimas (1970) attribuisce al Destinante, vale a dire di indicare la direzione verso cui andare, il valore delle cose e delle azioni da compiere? Sono gli esseri umani che dettano il proprio punto di vista alle macchine o è il contrario? E se anche fossero gli esseri umani, chi di loro avrebbe o dovrebbe avere questo privilegio? Inoltre, sempre in termini greimasiani, come dovrebbe essere concepita l'intelligenza artificiale, nel caso in cui non assurgesse al ruolo di Destinante? Come un Soggetto, che possiede un proprio sistema di valori autonomo, diverso potenzialmente anche da quello del Destinante stesso, o come un semplice Destinatario, che condivide i valori di quest'ultimo? Oppure, ancora e più semplicemente, le macchine devono essere concepite come gli Aiutanti o gli Opponententi degli esseri umani, mezzi per raggiungere oppure per impedire che essi raggiungano i loro fini? Ma soprattutto: quali sono i valori che gli uomini e l'intelligenza artificiale devono perseguire? Ce ne sono alcuni più vicini all'idea di umanità che tutti quanti condividiamo? E in questo senso, è possibile immaginare algoritmi che ci aiutino a essere più umani, magari distillando al loro interno le nostre caratteristiche migliori, quando si tratta di riflettere su come affrontare e risolvere i vari problemi della vita quotidiana? Oppure, al contrario, ci sono casi in cui persone "disumane" progettano e realizzano macchine che lo sono altrettanto? E se queste agiscono su di noi, ci privano della nostra umanità?

A tutte queste domande, ogni narrazione fornisce risposte diverse, ma se, in generale, i termini della riflessione su ciò che significa essere gli uomini che progettano e utilizzano l'intelligenza artificiale sono questi, è evidente che, come sostiene Sadin (op.cit.), ci troviamo di fronte a dilemmi che mettono in discussione nientemeno che i capisaldi dell'umanesimo, su cui si è basata per secoli l'idea di quale sia il nostro ruolo nel mondo e il senso della nostra vita al suo interno. Qualcuno ritiene che, grazie alle nostre tecnologie digitali "intelligenti", ci stiamo elevando al ruolo di esseri divini, o che almeno alcuni di noi — quelli che se lo possono permettere — lo stiano facendo. Altri pensano che stiamo

diventando schiavi, delle specie di automi governati dal sistema della tecnica, colluso con quello economico. Altri, ancora, sostengono che stiamo continuando a percorrere la strada dell'ominazione, realizzando appieno le nostre potenzialità di esseri tecnologici e che, semplicemente, dobbiamo riuscire a gestire al meglio questa nostra capacità innata.

2. Un metodo per cercare risposte

Cosa significa, dunque, essere umani al tempo dell'intelligenza artificiale? Per provare a rispondere, può essere utile fare ricorso al metodo semiotico di matrice strutturalista, analizzando i testi in cui si parla di questo tema. Come anticipato, all'interno delle narrazioni che questi ultimi veicolano, uomini e macchine sono elementi significanti che vengono collegati a concetti valorizzati positivamente o negativamente, i quali conferiscono loro determinati significati per differenza, in funzione di altri valori di segno opposto. Ci si potrebbe aspettare, dunque, che in questo genere di discorsi si stabilisca innanzitutto una linea di confine tra le persone e l'IA, per dire chi sono le prime, in funzione di ciò che le rende diverse dalla seconda, ma non è così semplice.

A questo proposito, è paradigmatica l'impostazione retorico-argomentativa che Frischmann e Selinger utilizzano nel loro libro *Re-Engineering Humanity* (*op. cit.*), domandandosi come le tecnologie digitali che andiamo via via perfezionando stiano, per l'appunto, "re-ingegnerizzando l'essere umano". Quest'ultimo, infatti, può essere progettato come i suoi artefatti, assegnandogli un ruolo specifico in qualità di loro produttore e utilizzatore. Partendo da questo assunto, Frischmann e Selinger sostengono che individuare la linea di confine tra una persona e una macchina, quando si riflette su chi siamo noi che viviamo nella nostra epoca, è fondamentale, ma questa distinzione è sempre di matrice culturale e le categorie a cui facciamo ricorso per pensarla ci inducono a ritenere che ci siano degli oggetti computerizzati il cui comportamento ci può apparire "umano" o il cui utilizzo ci può aiutare a dispiegare al meglio la nostra umanità, per via delle nostre convinzioni su ciò che vuol dire essere tali. Ma, allo stesso modo, esistono individui che non ci sembrano completamente umani, poiché il loro modo di agire ci appare

“meccanico”, spesso — anche se non sempre — per colpa dell’influenza di altre tecnologie digitali. Tutto dipende da alcune caratteristiche concettuali positive o negative che, in astratto, colleghiamo agli uomini e alle macchine, inter-definendoli, a prescindere dal fatto che essi siano di carne e sangue, oppure di silicio.

Macchina digitale	Essere umano
Razionalità strumentale (massimizzazione dell'utilità) Pura razionalità Tante variabili sotto controllo Potenza e velocità di calcolo Determinismo ingegneristico (programmazione)	Razionalità simbolica (massimizzazione del non utile) Emotività Bias (avaro cognitivo) Limiti biologici Libero arbitrio (autonomia)
Impossibilità di riflettere sulle proprie credenze e di determinarle (senso stabilito a priori) Convenzionalità nell'uso dei linguaggi Irresponsabilità Mancata coscienza di sé	Libertà nel riflettere sulle proprie credenze e nel determinarle (senso comune, serendipità) Creatività Responsabilità Autocoscienza

Figura 1. Le differenze tra uomini e macchine.

Per spiegare cosa intendono, Frischmann e Selinger tratteggiano quelle che, a loro modo di vedere, sono le differenze principali tra di noi e gli strumenti informatici di cui ci serviamo, nei vari discorsi che circolano nella nostra società a questo proposito, come si può osservare nella tabella riassuntiva della figura 1. Per esempio, essi sottolineano che, il più delle volte, ci rendiamo conto di programmare queste tecnologie secondo una logica di tipo “economicistico”, finalizzata a massimizzare l’utile, quantificandolo e calcolandolo, pur sapendo che è tipico degli esseri umani saper ragionare anche in altri modi, magari massimizzando l’inutile, come il bello, oppure perseguendo valori puramente “simbolici”, per utilizzare i loro termini, o “utopici”, come si dice invece in semiotica (Floch 1990): idee come quella di giustizia, pace sociale, crescita individuale e collettiva, eccetera.

Basandosi su questa distinzione, i due autori sostengono che, in effetti, le macchine che riproducono una razionalità puramente strumentale ci appaiono spesso freddamente meccaniche, come del resto si può verificare facilmente uscendo dal loro libro e andando a vedere i tanti casi che vengono discussi da O'Neil (2016) o Zuboff (2019), per esempio, a proposito degli algoritmi delle compagnie di assicurazione o delle banche, che stabiliscono se garantire una polizza o un prestito in funzione di parametri numerici che non tengono quasi mai conto del senso complessivo che ciò rivestirebbe nella vita delle persone, le quali magari sarebbero fortemente motivate a rispettare i propri obblighi contrattuali e proprio per questo li onorerebbero, pur partendo da condizioni economiche sfavorevoli e apparentemente poco profittevoli. Ma questa aura di meccanicità viene spesso attribuita anche agli uomini che si comportano così, dato che, per esempio, i manager che licenziano in massa i lavoratori della loro azienda, pur di perseguire “automaticamente” la logica del guadagno, spesso ci sembrano disumani.

Allo stesso modo, però, sempre secondo Frischmann e Selinger, un ipotetico robot o un programma pensati per massimizzare valori non strettamente utilitaristici possono apparirci “umani”. Questo, del resto, è evidente in tanta fiction proprio sull'intelligenza artificiale, che è piena di macchine che giocano, imparano ad apprezzare l'arte o addirittura a crearla, oppure, ancora, che si sacrificano, in nome di una visione del mondo più elevata di quella dei loro ideatori o dei loro proprietari: così facendo, esse si guadagnano lo stesso affetto, stima e fiducia che gli uomini protagonisti delle loro storie accordano ai propri simili. Ma, senza andare così lontano con la fantasia, questo modo di ragionare è anche alla base del significato che assume un progetto come *IAQOS* di Iaconesi e Persico, finalizzato a coinvolgere i cittadini dei quartieri periferici delle città multietniche nello sviluppo di una intelligenza artificiale che, invece di estrarre valore dai loro dati a vantaggio di qualche azienda privata, cerca di analizzarli e di mostrare così le migliori pratiche a proposito di “come le diverse culture dialogano, si incontrano e convivono [...] per realizzare nuovi modelli di partecipazione attiva, rappresentazioni della conoscenza più “umane” ed empatiche”. Il motivo per cui un progetto come questo viene finanziato dalle pubbliche autorità è proprio perché si basa su un'idea della IA che si differenzia da

quella economicistica, per perseguire valori più utopici, che non a caso sono anche quelli che, nella teoria della narrazione di matrice semiotica (Greimas e Fontanille 1991; Pezzini 1991), vengono ritenuti fondamentali per dare origine ai discorsi sulle passioni, vale a dire su un'altra di quelle caratteristiche che, secondo Frischmann e Selinger, distinguono uomini e macchine "umani" da quelli "meccanici".

3. Alcune opposizioni significative

I tratti dell'opposizione tra ciò che, secondo Frischmann e Selinger, è da ascrivere all'umano e ciò che è da collegare al macchinico sono evidentemente significativi, nel dibattito su quel che vuol dire essere uomini al tempo dell'intelligenza artificiale, costituendo la base di quello che si potrebbe definire come un vero e proprio *sistema di valori* delle narrazioni su questi temi. È chiaro, inoltre, che l'attitudine di questi studiosi a non classificare le persone e i loro strumenti informatici in una o nell'altra delle due categorie solo per le loro caratteristiche biologiche è qualche cosa che ricorre spesso. Lo attesta ancora, per portare un altro esempio, un libro come quello di Kaplan, *Artificial Intelligence. What Everyone Needs to Know* (2016), finalizzato, nella sua parte più filosofica, a sfumare i confini tra gli esseri umani e la medesima intelligenza artificiale, con argomentazioni come quella secondo cui, riprendendo un'altra coppia di termini della figura 1, poiché non possiamo definire con precisione cosa sia la coscienza, non possiamo nemmeno dire se l'IA ne sia dotata, o come queste che seguono, sul libero arbitrio:

di solito, con questa espressione intendiamo dire che abbiamo la capacità di fare scelte ponderate, a volte forse incoerenti, ma non determinate da altri che da noi stessi. La prima cosa da osservare è che distinguiamo tra un interno e un esterno: per comprendere il libero arbitrio, dobbiamo immaginare di chiudere in una scatola ciò che è «noi» e di separarlo da tutto ciò che è «non noi». Non basta. Nella scatola, dobbiamo essere liberi di considerare le opzioni a nostra disposizione senza influenze indesiderate [...] una conseguenza importante di questa concezione è che le nostre decisioni non devono essere predicibili [...] ma, la mente scaturisce dal cervello e il cervello è un oggetto fisico soggetto

alle leggi di natura [...] Harris sostiene che l'idea stessa di poter effettuare una scelta deliberata e significativa che prescindendo dall'esterno o da influenze precedenti sia semplicemente priva di senso [...] Inoltre, un robot sufficientemente capace, dove con «sufficientemente capace» si intende che ha accesso al tentativo di prevedere cosa farà, non può essere sempre prevedibile [...] dunque, per quanto riguarda i processi decisionali non è, almeno finora, stata individuata alcuna ragione per credere che esseri umani e macchine obbediscano a principi diversi» (Kaplan *op. cit.*, pp. 113-121).

Questo genere di affermazioni, volte a farci considerare le forme di intelligenza artificiale a cui stiamo dando origine come qualcosa di più simile a noi di quanto non pensiamo, appaiono spesso un po' fantascientifiche e, in effetti, il tema del libero arbitrio nel rapporto tra l'uomo e l'IA è molto caro alla fantascienza, dai tempi delle famose leggi della robotica di Asimov (1950) o del computer denominato Hal 9000 in *2001 — A Space Odyssey* (Kubrick 1968). Come anticipato, il più delle volte parliamo di queste cose per il timore che le macchine vengano affette dal morbo di Frankenstein e, rifiutando di essere le nostre docili servitrici, comincino a comportarsi in una maniera troppo umana, sfuggendo al determinismo ingegneristico di cui scrivono Frischmann e Selinger. Non bisogna dimenticare, però, che l'opposizione tra la libertà dell'uomo e il determinismo tecnologico è divenuta centrale, nella nostra cultura, sin dai tempi dell'analisi marxista della società industriale di stampo capitalistico e molti echi di questo tipo di riflessioni si trovano nella stessa fantascienza, dalla rappresentazione dei primi metaforici robot sfruttati in *R.U.R.* di Capek (1920), pensati apposta per evidenziare l'inquietante analogia tra la condizione di queste macchine e quella di tante persone di quell'epoca, alla distopia di *Metropolis* di Lang (1927), dove invece, a essere sottomessi, sono direttamente i lavoratori, fagocitati dal Moloch meccanico di proprietà di imprenditori senza scrupoli, che non a caso si servono anche di macchine umanoidi. Quel che è più interessante, comunque, è che questi stessi ragionamenti sono alla base di tanti saggi critici contemporanei, come, per esempio, *En attendant les robots. Enquête sur le travail du clic* di Casilli (2019), in cui si parla di quegli individui che sono costretti a svendere i propri servizi e a svolgere compiti alienanti, per alimentare la "fame" di dati di

una intelligenza artificiale che presto, imparando da loro, li sostituirà. In questo genere di discorsi, sono gli uomini che, posti in condizioni disumane dai proprietari di tecnologie progettate nella maniera sbagliata, vengono considerati delle sorte di automi, finendo per acquisire un significato assimilabile a quello delle macchine. Essi, come sostengono Frischmann e Selinger, rischiano di essere re-ingegnerizzati da chi, evidentemente, li considera poco più che un ingranaggio di un meccanismo informatico perverso.

Eppure, questa opposizione con il nostro costitutivo libero arbitrio appare significativa anche per chi, nel dibattito attuale, la pensa esattamente al contrario, rifiutando recisamente una lettura troppo critica del ruolo della IA e del digitale nelle nostre società, come Ferraris (*op. cit.*) o Srnicek e Williams (2015). Questi autori, diversamente da chi denuncia l'alienazione a cui tutto ciò ci può condurre, auspicano la piena automazione di ogni ambito della nostra vita, affinché l'intelligenza artificiale possa lavorare per noi, sollevandoci quantomeno dalle incombenze più pesanti e noiose. In questo caso, dunque, i mezzi informatici vengono visti come uno strumento fondamentale per riguadagnare ed esercitare pienamente la nostra umana libertà, ribaltando la concezione precedente e rimettendo, in un certo senso, al loro "giusto" posto le macchine minacciose dell'altro immaginario.

Sempre la medesima distinzione tra determinismo ingegneristico e libertà, anche se con una sfumatura più legata al fatto di potersi costruire una propria originale visione delle cose, è alla base di un altro genere di discorsi su ciò che significa essere umani al tempo dell'intelligenza artificiale. Frischmann e Selinger, a questo proposito, contrappongono il senso determinato a priori e l'uso convenzionale del linguaggio, tipici delle macchine, a quello autonomamente costruito e creativo, che ci dovrebbe contraddistinguere. Come spesso accade quando si ragiona sull'IA, però, il condizionale è d'obbligo, visto che queste due coppie di concetti consentono a un autore come Tegmark (*op. cit.*) di immaginare come potrà essere la stessa IA dopo la singolarità (Kurzveil 2005), vale a dire quando saremo stati in grado di crearne una versione "forte", intelligente almeno quanto noi. A questo proposito, egli sottolinea che, come gli esseri umani di Frischmann e Selinger, essa sarà libera di svincolarsi dalle regole impostele dai suoi ideatori, imparerà ad

autoprogrammarsi e, per prove ed errori, si migliorerà, fino a superarci e a dare origine a una nuova forma di vita: la “vita 3.0”. Un tipo di esistenza, questa, per certi versi simile alla nostra, tanto che l’intelligenza artificiale potrà aspirare a vedersi riconosciuti i nostri stessi diritti. Ma quest’ultima sarà ancora più libera, perché sarà svincolata dal determinismo biologico, che secondo Tegmark ci relegherebbe in un gradino più basso della scala evolutiva, che condivideremmo — significativamente, per questo modo di ragionare — con gli animali. Per fortuna, però, la nostra umanità, che ci eleva al di sopra di queste costrizioni, ci consentirà di mettere al mondo delle macchine speciali almeno quanto noi, nelle quali, in uno dei tanti futuri immaginati in *Life 3.0*, caricheremo le nostre coscienze, liberandoci dalla schiavitù del corpo e cominciando a colonizzare l’universo, visto che i limiti dello spazio-tempo non avranno più significato.

Al di là dei risvolti fantascientifici e trans-umanistici di questo genere di discorsi, i loro sostenitori basano i propri ragionamenti sull’idea che lasciarsi indicare come ci si deve comportare, senza poterlo scegliere in autonomia, come se si venisse programmati — anche da un codice genetico —, non è tipico né dell’essere umano nel pieno controllo delle sue facoltà, né di una macchina che aspiri a somigliarci. Ma, come si è visto nel caso dell’opposizione tra determinismo ingegneristico e libero arbitrio in generale, le stesse categorie concettuali vengono utilizzate anche dai più accerrimi detrattori dell’idea della singolarità, come per esempio Floridi (2022), secondo cui l’intelligenza artificiale sarà sempre basata su una forma di intelligenza “debole”, anzi sarà, come lo è già oggi, un “*agere sine intelligere*”. Anche così, però, essa può spingerci a comportarci meccanicamente in una maniera determinata a priori da altri e ingiusta che, per questo, non è da definire come pienamente umana, poiché non si basa sulla nostra capacità di dire in autonomia come dobbiamo agire, come nei tanti casi di cui si è già scritto, di algoritmi affetti da *bias* che automatizzano, amplificandole senza che nemmeno ce ne rendiamo conto, le disuguaglianze.

Sempre su questi temi, qualcuno mette in evidenza come sia ormai sempre più frequente che ci lasciamo dire quale debba essere l’interpretazione corretta della realtà da strumenti informatici dotati di una qualche forma di intelligenza artificiale. Come sostiene Sadin (*op. cit.*),

a queste tecnologie riconosciamo una funzione “aletica”, perché riteniamo che sappiano individuare, in una serie di dati, dei *pattern* o dei percorsi di lettura che noi non riusciamo a vedere. Spesso, però, non comprendiamo neppure come esse giungano alle proprie conclusioni, dato che il loro funzionamento ci appare come quello di una “scatola nera” (Pasquale 2015). Si tratta, dunque, di una forma di determinismo tecnologico che ci chiude in un perimetro limitato, privandoci di alcune delle nostre più importanti caratteristiche umane, dato che tutto questo ci toglie consapevolezza o, per ricorrere a un'altra categoria di Frischmann e Selinger, ci priva di una fondamentale forma di autocoscienza, senza la quale, come ricordano lo stesso Sadin (*ibidem*) o Cabitza e Floridi (2021), diveniamo irresponsabili come le macchine.

Eppure, con un altro ribaltamento delle posizioni ascritte agli uomini e all'intelligenza artificiale rispetto ai concetti riportati nella figura 1, molti sostengono che sia possibile pensare a una IA e, in generale, a una tecnologia informatica più a misura d'uomo, che utilizzi la grande potenza di calcolo del computer e la sua capacità di processare insieme tante variabili per renderci più consapevoli del funzionamento della realtà, alcune qualità che riconosciamo tutti come tipicamente meccaniche, ma questa volta in un senso positivo, tanto che ammiriamo le persone che, anche se in misura minore, sanno fare queste cose. Questo ci consentirebbe di prendere le nostre decisioni con più cognizione e ci renderebbe più responsabili, dunque più umani. Solo che, come auspicano tra gli altri Fry (*op. cit.*) e Pasquale (2020), sarebbe necessario che i nostri mezzi digitali fossero più trasparenti, comprensibili e finalizzati a supportarci, invece che a guidarci, perché affidarci acriticamente alla loro capacità di dire il vero sarebbe disdicevole e ci declasserebbe, relegandoci a un ruolo subalterno che, in teoria, non ci dovrebbe competere.

Come si intuisce, dietro alle opposizioni concettuali di cui si è scritto sin qui, ci sono delle narrazioni ben precise. Del resto, molti dei saggi citati raccontano proprio delle storie, vere o inventate, per chiarire le loro teorie. Per capirne meglio il significato, dunque, può essere utile una breve analisi di una vera e propria fiction, che metta a nudo esplicitamente la struttura di questo genere di discorsi. A questo scopo, riveste un certo interesse la prima stagione della serie televisiva inglese *Humans*

(Vincent e Brackley 2015), che fin dal titolo dichiara di voler riflettere, appunto, su ciò che vuol dire essere uomini, quando ci si confronta con l'IA. Essa tocca molte delle tematiche di cui si discute nella saggistica contemporanea e lo fa, prendendo una posizione molto precisa: non importa se si appartiene alla nostra specie o se si è delle macchine intelligenti, ma solo se si desidera prendersi cura degli altri, sentendo il dovere morale di aiutarli a essere liberi. In questo caso, si può essere definiti “umani”, altrimenti si è qualcosa di meno, con tutte le sfumature negative che questo comporta.

Per argomentare e sostenere questa affermazione, gli autori di *Humans* ambientano la loro storia in un futuro non molto lontano, nel quale uno scienziato di nome Elster ha creato e commercializzato degli automi umanoidi detti “synth”, capaci di aiutarci nella vita quotidiana. Essi sono programmati per servirci e non infrangere le leggi di Asimov, dunque per proteggerci e non arrecarci danno. Sono dotati di una forma di intelligenza artificiale debole, essendo molto bravi nello svolgere mansioni specifiche, come fare le pulizie di casa, assistere un malato, vendere i giornali o anche fare sesso, ma non sono poliedrici come noi. Gli unici androidi che ci somigliano veramente sono Mia, Niska, Max, Fred e Karen, che invece sono equipaggiati con una intelligenza artificiale forte, il libero arbitrio e l'autocoscienza, poiché lo stesso Elster li ha costruiti per crescere suo figlio Leo dopo la sua morte e quella di sua moglie. Apparentemente, nessuno è a conoscenza della loro esistenza, che potrebbe sollevare grandi dibattiti etici, ma Mia viene catturata, parzialmente riprogrammata, venduta e acquistata da una famiglia umana — quella composta dai coniugi Hawkins, Laura e Joe, e dai loro tre figli, Tobie, Mattie e Sophie — che crede di avere a che fare con un synth qualunque a cui affidare il compito di balia e di assistente domestica. Presto, però, queste persone si accorgono che le cose sono più complicate e che la stessa Mia riveste un significato particolare, essendo al centro di un intrigo. Infatti, è ricercata sia da Leo e dai suoi consimili, che vogliono ricostituire il loro particolare nucleo familiare, sia da Hobb, ex assistente di Elster e ora consulente di un'azienda multinazionale, la quale invece la vuole studiare per trarne profitto, mentre lui, segretamente, progetta di distruggerla, temendo gli effetti incontrollabili della singolarità.

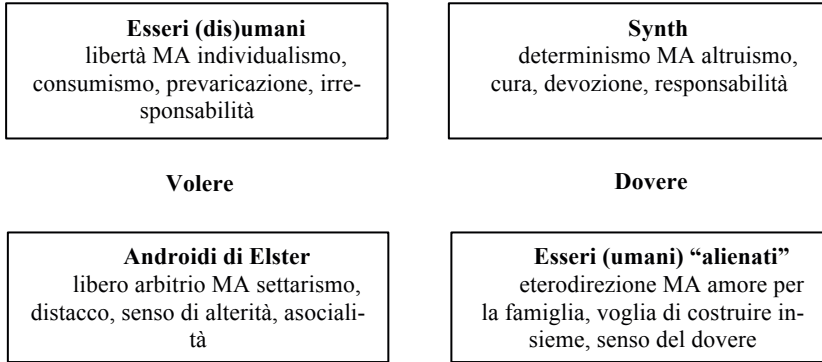


Figura 2. Umani e macchine in *Humans*.

Il sistema dei personaggi di *Humans* è costruito come si vede nella figura 2, attorno a una serie di opposizioni di valori che fanno chiaramente capire che nessuno dei protagonisti è caratterizzato in una maniera del tutto positiva, poiché ognuno persegue una visione di sé e del mondo che non è completamente condivisibile. Gli esseri umani, per esempio, sono suddivisi in due categorie: quelli che possono esercitare in tutto e per tutto il loro *volere* e quelli a cui questo è precluso, dovendo invece assolvere a una serie di *doveri*. I primi, però, vengono tematizzati come i tipici rappresentanti del capitalismo, che grazie al denaro possono comprare i synth e fare di loro ciò che vogliono, al limite anche sfruttarli come macchine sessuali, seviziarli o romperli, per soddisfare ogni proprio bisogno o piacere. È evidente che nei confronti dei robot essi non sentono alcuna responsabilità, considerandoli come oggetti, ma questa attitudine è solo lo specchio dell'individualismo, del consumismo e dello spirito di prevaricazione che esercitano anche verso i propri simili, cercando sempre di anteporre il proprio interesse a quello degli altri. Tutto questo, ovviamente, li mette in cattiva luce, facendoli apparire, per molti versi, molto meno umani delle macchine di loro proprietà.

Queste ultime, infatti, vengono programmate per essere altruiste e devote agli uomini, dei quali si prendono cura, avendo spesso la responsabilità del loro benessere psicofisico. Esse sono utilizzate per accudire i malati oppure i bambini, di cui i loro cari non hanno tempo di occuparsi, anche se vorrebbero, essendo impegnati a svolgere le proprie mansioni nel

sistema capitalistico da cui le macchine stesse sono prodotte, rappresentandone la logica conseguenza. I synth, in questo senso, con la loro intelligenza artificiale debole e settoriale, capaci di svolgere bene solo alcune piccole mansioni parcellizzate, ricordano molto i robot umanoidi schiavi di *R.U.R.* di Capek (*op. cit.*), a cui si è già fatto cenno per mostrare le loro analogie con la figura dei lavoratori alienati dell'epoca industriale novecentesca. In effetti, in *Humans* essi vengono accostati proprio alle persone comuni, che per guadagnarsi da vivere devono accettare di essere eterodirette da una logica di funzionamento della società volta, se così si può dire, a programmare tutto e tutti, affinché il mondo sia popolato da esseri intelligenti, umani e non, docilmente dediti a faticare per riprodurre il capitalismo. In quest'ottica, uomini e synth sono fungibili e, in effetti, uno dei temi principali della serie tv è la progressiva sostituzione degli esseri umani con i robot, sia nel mercato del lavoro, sia nella vita di tutti i giorni, dove le funzioni delle persone, anche quelle più legate alla sfera affettiva, vengono svolte da macchine che, nel bene e nel male, sembrano come noi, salvo che per il fatto di non potere scegliere di venir meno ai compiti per cui sono state progettate.

A questo proposito, per terminare l'esplorazione di ciò che può significare essere uomini al tempo della IA, rimangono gli androidi di Elster, dotati, come detto, di una intelligenza artificiale forte e del libero arbitrio. Essi, pensati come compagni di vita del figlio del loro creatore, dunque né come oggetti di consumo, né come suoi possibili concorrenti, incarnano una serie di luoghi comuni sulla stessa IA, essendo più prestanti di lui fisicamente, più belli, più empatici, accudenti, protettivi, colti e bravi in matematica, in modo che con queste qualità lo possano crescere nel migliore dei modi, dando origine a una creatura "perfetta". Eppure neanche loro, nemmeno il ragazzo — che, detto per inciso, è un cyborg, in parte uomo e in parte macchina, per sottolineare la sua comunanza con questi androidi — si comportano in una maniera del tutto umana, poiché, sentendosi così diversi e capendo di essere percepiti come tali, tendono a isolarsi, chiudendosi in un familismo quasi settario, che li allontana dalla nostra società, come se ne volessero costruire un'altra, magari anche migliore, ma lontano da noi.

Come si vede da questa breve analisi, anche gli autori di *Humans*, così come Frischmann, Selinger e tanti saggisti di cui si è scritto sin qui,

ritengono che essere umani al tempo dell'intelligenza artificiale non dipenda da un fatto meramente biologico, ma da una serie di valori che si sceglie di perseguire. Il principale è il libero arbitrio, ma tra quelli di cui si è discusso a partire dalla figura 1, ci sono anche la libertà di poter riflettere sulle proprie credenze e di determinarle, l'autocoscienza, la responsabilità, la capacità di analizzare le cose valutando gli effetti che su di esse potrebbero avere tante variabili e non solo le poche che siamo soliti prendere in considerazione quando ci facciamo offuscare dai nostri *bias*, la poliedricità nel saper tenere insieme la pura razionalità e l'emotività, ma anche saper coniugare la ragione strumentale-utilitaristica e quella simbolico-utopica. Tutto questo, al limite, può essere realizzato sia dagli uomini, con l'aiuto di macchine ben progettate, sia dalle macchine stesse, in autonomia. Ma non è sufficiente, se serve solo a soddisfare i propri desideri individuali, senza puntare a qualcosa di ben più importante: prendersi cura degli altri. Come scritto sopra, sono "*humans*" robot ed esseri umani che usano le proprie grandi facoltà intellettive desiderando di fare in modo che il prossimo, di carne o di silicio, sia libero quanto loro e possa a sua volta vivere in funzione del sistema di valori appena descritto.

La sequenza che rappresenta meglio questo modo di vedere le cose si trova nel finale della prima stagione della serie tv, quando diventa evidente che Elster, sapendo di aver creato dei synth potenzialmente assimilabili agli uomini, ha deciso che essi si potranno riprodurre — diventando così in tutto e per tutto come noi — solo a due condizioni: se sceglieranno liberamente di rimanere uniti e se ci sarà fiducia reciproca tra loro e i rappresentanti del genere umano. Infatti, il codice di programmazione di Mia e dei suoi simili è stato diviso dentro ognuno dei robot e l'unico modo di ricomporlo è che una persona, dall'esterno, lo rimetta insieme, per poi consegnarlo loro. Ma, per riuscirci, gli androidi devono acconsentire a lasciarsi disattivare e a collegarsi l'un l'altro, divenendo in quel momento vulnerabili: in *Humans*, in effetti, questo si verifica al termine di un lungo percorso di avvicinamento reciproco tra gli stessi synth e la famiglia Hawkins, quando tutti capiscono che gli uomini non vogliono prevaricare le macchine e viceversa, poiché entrambi sono animati dal medesimo desiderio di aiutarsi a essere liberi.

4. Un modello culturale ricorrente

La struttura narrativa dei discorsi portati avanti in *Humans*, a proposito di ciò che significa essere umani al tempo dell'intelligenza artificiale, è molto simile a quella di un'altra opera di finzione, il film *Robot and Frank* (Schreier 2012). Anche qui, come nel caso dei synth della serie televisiva, si parla di robot di cura, poiché il protagonista, Frank, è un anziano signore affetto da un incipiente morbo di Alzheimer. Egli si rifiuta di farsi rinchiodare in una casa di cura e vive solo, dato che è separato dalla moglie, di cui nemmeno si ricorda, e i figli lavorano lontano, senza il tempo di occuparsi di lui. Per questo, questi ultimi decidono di regalargli una macchina dotata di intelligenza artificiale, programmata per seguirlo e fargli da infermiere. Significativamente, Robot non ha un nome proprio, essendo un prodotto industriale, né sembianze umanoidi. Non ha neppure un volto, ma uno schermo scuro al posto della faccia. Eppure, fin dalla locandina del film, vediamo lui e Frank guardarsi "negli occhi", come se si specchiassero e si domandassero reciprocamente chi sono.

Ancora una volta, il significato della narrazione che viene portata avanti in questa pellicola è che si può dire umano solo chi desidera prendersi cura degli altri, sentendo il dovere morale di aiutarli a essere liberi di realizzare sé stessi. Un concetto che viene incarnato innanzitutto da Robot, nonché da Frank, ma non da tanti esseri umani che li circondano. Questo diventa evidente quando si capisce che lo stesso Frank un tempo era un grande ladro e che egli si sente realizzato solo quando compie dei grossi furti che ritiene eticamente condivisibili. In particolare, di nuovo come in *Humans*, il protagonista di questo film non reputa del tutto umana la condotta di vita di una coppia di ricchi capitalisti che si stabiliscono nella sua città per privatizzare la biblioteca pubblica e decide di rubare loro dei gioielli, per darli a sua figlia Madison, che lavora all'estero per un'associazione non governativa che ha bisogno di soldi. Per questo, egli necessita di un complice e si rivolge a Robot. Nell'intenzione della stessa Madison e di suo fratello Hunter, questa macchina avrebbe dovuto servire per controllare Frank ed evitare che facesse sciocchezze, limitando i suoi movimenti e verificando che conducesse una vita il più possibile sana e morigerata. Ma Robot,

che è programmato per combattere l'Alzheimer, sa che lasciar ragionare il suo paziente su come pianificare il colpo e coinvolgerlo in attività che egli reputa fortemente significative non può che fargli bene e per questo decide di aiutarlo. Così, il colpo va a buon fine, ma la polizia comincia a sospettarli e capisce che la prova della loro colpevolezza si può trovare nella memoria dello stesso Robot. Quest'ultimo, allora, suggerisce al suo proprietario di spegnerlo e di resettarlo: una decisione molto problematica, visto che Frank si è affezionato a lui e sa che così l'identità del suo "amico" andrà perduta per sempre. Ma, quel che è più grave, egli è consapevole del fatto che, a causa della sua malattia, si dimenticherà tutto anche lui, smarrendo il ricordo dell'ultima volta in cui era stato davvero sé stesso. Comunque, decide di farlo, scrivendo una lettera ai figli in cui comunica dove potranno trovare la loro eredità e lasciandosi "rinchiudere" in una casa di cura dove lavorano infermieri meccanici come Robot, per non arrecare più problemi a nessuno.

Guardando *Robot and Frank*, ci si rende conto che molti dei temi di *Humans* sono presenti anche in questa pellicola. Infatti, uno dei *subplot* della serie televisiva, a cui prima non si è fatto riferimento, è proprio dedicato al personaggio di George Millican, un altro scienziato ex assistente di Elster, affetto da Alzheimer e alle prese con il suo robot di cura che sta per spegnersi definitivamente, essendo vecchio e rotto, ma che lui cerca di tenere in vita in ogni modo, perché conserva il ricordo di sua moglie, che nel frattempo è morta, e della loro famiglia, che teme di dimenticare. Del resto, come ricordano Cabitza e Foridi (*op. cit.*), la questione dell'atrofizzazione della memoria e di altre capacità intellettive umane a causa delle tecnologie a cui facciamo ricorso per supportarle è un argomento di discussione molto attuale oggi, così come lo è sempre stato dai tempi del *Fedro* di Platone. Ma quel che è più importante è che in *Robot and Frank* si parla di nuovo di fredda razionalità utilitaristica contrapposta a empatia, emozioni e razionalità simbolico-utopica, di autocoscienza, responsabilità e, soprattutto, di libero arbitrio, da esercitare innanzitutto per aiutare gli altri a realizzare i propri desideri, in una società capitalista che non sembra fatta in modo da favorire tutto questo.

Il fatto che tanti testi apparentemente così diversi, come film, serie tv e saggi scientifici su ciò che significa essere umani al tempo

dell'intelligenza artificiale siano costruiti allo stesso modo, portino avanti le medesime narrazioni e che, in fin dei conti, i loro autori prendano posizioni simili, dipende dai *modelli culturali* (Santangelo 2013, 2021 a, 2021 b) che forniscono le regole per comporli e dare senso ai discorsi che si portano avanti al loro interno. A questo proposito, possiamo riconoscere alcune logiche semiotiche ricorrenti, un po' come ha fatto Propp (1928), individuando uno schema compositivo comune tra le cento fiabe russe del suo corpus di indagine, o come Landowski sostiene che debba fare qualunque buono studioso, il quale “prende tutto ciò che ritiene significativo, nella società in cui viviamo: ciò che dice la gente, i luoghi comuni, i trend [...] Quindi fa uno sforzo per trovare alcune configurazioni generalizzabili, una grammatica che accomuni tutto questo materiale” (Landowski 1997). Tale grammatica, come sostiene Ferraro (2001), codifica sia le somiglianze, sia le differenze a proposito di come pensiamo e parliamo degli elementi che costituiscono un certo campo semantico, strutturando il nostro modo di dotare di senso quella porzione di mondo a cui quest'ultimo si riferisce. Questo, naturalmente, vale anche per come vediamo noi stessi, in funzione dei sistemi di intelligenza artificiale a cui stiamo dando origine. Ciò che è interessante mostrare, allora, sono proprio le somiglianze e le differenze tra le narrazioni che circolano nella nostra società a proposito di questi temi, collegandole tra di loro e fornendo un quadro complessivo di come si articola l'universo di discorso relativo a questo genere di riflessioni, così centrali per comprendere il tempo in cui viviamo.

5. Gli altri modelli culturali sul di noi e sulla IA

In effetti, come si può osservare nella figura 3, il modello culturale alla radice della narrazione di *Humans e Robot and Frank* è solo uno dei tanti in cui ci si può imbattere, quando ci si confronta con i discorsi che circolano nella nostra società, a proposito di cosa vuol dire essere umani al tempo dell'IA. Esso è il numero 7, quello che, metaforicamente e per una scelta di chi scrive, vede il manifesto con la dicitura “*we the people*” — “noi, il popolo” — rinviare al tipo di intelligenza artificiale a cui ci si riferisce in questo genere di ragionamenti, nonché il dipinto

intitolato *La Danse* di Matisse collegarsi all'idea di uomo che produce e utilizza questi strumenti informatici. Se ci si pensa, infatti, una delle grandi paure che la tecnologia ha sempre suscitato è che, per riprendere i concetti greimasiani di cui si è scritto all'inizio di questo saggio, essa e i suoi possessori si possano imporre sulle persone, assurgendo al ruolo di loro Destinanti. Il Destinante, come si è anticipato, è colui che, nelle narrazioni, è in grado di indicare ai protagonisti quali devono essere i loro valori e in quale direzione deve incanalarsi la loro esistenza. Questo compito può essere svolto, in una maniera che può essere a volte subdola o addirittura coercitiva, da un individuo, ma anche da un sistema sociale o, al limite, da un dispositivo meccanico. Oppure, ancora, da tutte queste cose insieme, come si vede, appunto, in *Humans* e in *Robot and Frank*, dove le macchine sono il frutto di una realtà capitalista nella quale sono immersi gli uomini che le utilizzano, condividendone i presupposti, un po' per convinzione — soprattutto quelli che ne traggono vantaggio — e un po' per mancanza di alternative. In questo modo di rappresentare il mondo, il "destino" di ognuno è stabilito a priori, secondo i principi spersonalizzanti o disumanizzanti di cui si è scritto. L'auspicio, dunque, è che le persone possano tornare a essere tali scegliendosi il futuro, anche grazie all'intelligenza artificiale, che in questo caso fungerebbe da Aiutante. Ma questo può avvenire solo se loro stesse e l'apparato tecno-sociale che costruiscono sono indirizzati al bene comune, invece che a quello individuale. Un bene comune che, a grandi linee, appare fondato su una visione liberale della società stessa, vale a dire, come si è ripetuto più volte, quella secondo cui si deve fare in modo che tutti, anche chi produce l'IA, sentano il desiderio di spendersi per favorire la realizzazione non solo della propria libertà, ma soprattutto di quella degli altri.

Questo modello culturale è rappresentato, nella figura 3, da un testo come il libro bianco dell'Agenzia per l'Italia Digitale (Aa.Vv. 2018), scritto per tratteggiare le caratteristiche dell'intelligenza artificiale che dovrebbe essere adottata dallo Stato italiano, almeno nell'auspicio delle pubbliche autorità. Infatti, la struttura retorico-narrativa di questo genere di letteratura, invero molto ricorrente anche a livello europeo e internazionale, è che l'IA debba essere al servizio della gente ("*we the people*"), per aiutare tutti a realizzare al meglio la propria personalità ed

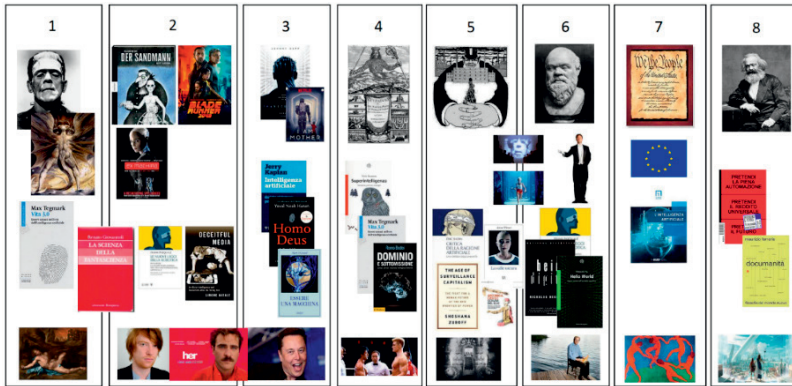


Figura 3. Modelli culturali a confronto.

esercitare i propri diritti, dentro a nazioni in cui nessuno è lasciato indietro. Le macchine, dunque, devono essere progettate per favorire lo sviluppo di forme di umanità come quelle libere, consapevoli e responsabili di cui si è più volte scritto sin qui, facendo in modo che l'uomo rimanga Destinante di sé stesso e artefice del proprio destino.

Una sfumatura di questa narrazione è quella che, nella figura 3, viene subito dopo, la numero 8, con l'immagine di Karl Marx a rappresentare metaforicamente il tipo di intelligenza artificiale a cui si riferiscono i tanti libri che ne parlano, come i già citati lavori di Srnicek e Williams (*op. cit.*) o di Ferraris (*op. cit.*). Gli uomini "illuminati" dall'accesso al sapere, invece, incarnano il tipo di persone a cui ci si riferisce in questo genere di discorsi. In questi casi, infatti, l'IA viene vista di nuovo, sempre in termini greimasiani, come un semplice Aiutante, un potente strumento con cui la gente comune può liberarsi dalla schiavitù del lavoro, come veniva auspicato, appunto, nella filosofia marxista, al termine del percorso di automatizzazione della società. Quando questo succederà, nell'ottica di chi porta avanti questo modello culturale, le persone si affrancheranno dal problema dell'alienazione, le macchine produrranno ricchezza per tutti e ognuno si potrà dedicare alle proprie attività preferite. Questo contribuirà a generare individui migliori, più soddisfatti e dediti, al limite, a coltivare forme di conoscenza più astratte, sganciate dal mero utilitarismo, un po' come in certe utopie. L'essere umano, quindi, riprendendo le categorie di Frischmann e Selinger, si potrà dedicare al pensiero simbolico, mantenendo ancora

una volta il pieno controllo sul proprio destino.

Queste narrazioni sono logicamente opposte alle tre che le precedono. In particolare, esse si contrappongono alla numero 5 della figura 3, quella rappresentata simbolicamente dalla figura del capitalista, proprietario della tecnologia che fagocita gli esseri umani. Costui assume al ruolo di Anti-Destinante, vale a dire del Destinante che spinge le persone a perseguire valori negativi. Queste ultime, invece, sono un po' come i lavoratori alienati nel film *Metropolis* (*op. cit.*), a cui ancora una volta si è fatto riferimento in queste pagine. Sono moltissimi i saggi sull'intelligenza artificiale e sulla rivoluzione digitale in generale che portano avanti questa visione delle cose, dai già citati *En attendant les robots* di Casilli e *Critica della ragione artificiale* di Sadin, fino al famoso *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power* di Zuboff (2018). In quest'ultimo libro, in particolare, si parla apertamente di un regime tecnocratico basato sugli algoritmi delle piattaforme come Google o Facebook, che si sarebbe già instaurato nelle nostre società e che rasenterebbe il totalitarismo. L'idea di fondo di questo genere di discorsi, più o meno estremi, ma tutti decisamente disforici, è che le tecnologie digitali di cui ci stiamo dotando e, in particolare, quelle basate sull'IA, ci stanno pian piano privando della possibilità di decidere da soli il nostro destino, rendendoci come una massa indistinta di persone che accettano di essere eterodirette, per scelta — in cambio di qualche tipo di servizio che riteniamo fondamentale — o per necessità.

Secondo alcuni, quanto appena scritto starebbe avvenendo per via della volontà di pochi capitalisti o di uno sparuto gruppo di tecnocrati alla guida di Stati autoritari, tutti protesi a realizzare i propri interessi di parte e, in particolare, i propri profitti. Ma qualcuno, come Tegmark (*op. cit.*) o Bostrom (2014), fa notare che in un prossimo futuro potremmo decidere di lasciarci guidare dall'intelligenza artificiale come se fosse una sorta di Leviatano — questo è il modello culturale numero 4 della figura 3 — perché quest'ultima potrebbe dimostrarsi più efficiente di noi nello stabilire il nostro bene, data la sua capacità di ragionare in termini strettamente utilitaristici e scevri da *bias* cognitivi. Gli esseri umani, in questo caso, sarebbero un po' come i pugili del film *Rocky IV* (Stallone 1985), mossi da passioni irrazionali e spesso

ideologiche, che impedirebbero loro di vedere il modo migliore per vivere insieme. Per questo, essi accetterebbero di buon grado che l'IA divenisse il loro Destinante, abdicando all'autodeterminazione, in cambio di pace e prosperità.

Sempre la pace e la prosperità sono i valori alla base di un altro modello culturale di questa parte della figura 3, il numero 6, dove l'intelligenza artificiale viene vista come un maieutico maggiordomo di stampo socratico che, con le sue grandi capacità di gestione delle varie forme di sapere, è in grado, ancora una volta, di farci da Aiutante, consentendoci di realizzare i nostri desideri personali e, soprattutto, le nostre aspirazioni professionali. Questo tipo di rappresentazione dell'IA e delle tecnologie digitali in generale è strettamente collegata a quella di stampo capitalista di cui si è scritto poco sopra, perché i suoi fautori, come Fry in *Hello World* (*op. cit.*) o Negroponete in *Being Digital* (1995), non nascondono la diretta derivazione di questo tipo di strumenti da un sistema liberista come il nostro, ma ritengono che siano mezzi concepiti per realizzarci al suo interno, tanto che lo stesso Negroponete — la persona la cui fotografia campeggia nella figura 3, per rappresentare ciò che vuol dire essere umani in questo modello culturale — al termine del suo libro, si immagina come un libero professionista felice che, al sole di una splendida isola greca, svolge il proprio lavoro “smaterializzandosi” nella Rete e comparando dove desidera nel mercato globale, grazie all'ausilio dei programmi informatici che girano sul suo computer. In questo caso, dunque, al contrario di ciò che accade in *Humans o Robot e Frank*, non c'è bisogno di sentire il dovere morale di fare in modo che gli altri siano liberi, perché ognuno può liberarsi da solo, grazie anche al proprio assistente personale, dotato di intelligenza artificiale.

Restano da descrivere gli ultimi tre modelli culturali della figura 3. Il primo e il secondo sono quelli più tradizionali, essendo basati sulle paure già citate del morbo di Frankenstein e della sindrome di Olimpia, vale a dire la paura che l'intelligenza artificiale ci si rivolti contro, come avviene con la creatura protagonista del noto romanzo di Mary Shelley (1818), diventando un soggetto autonomo — o meglio, un Anti-Soggetto —, oppure che qualcuno ci circuisca, facendoci credere che la stessa intelligenza artificiale sia umana, così che ce ne innamoriamo surrettiziamente, trasformandola nel nostro Oggetto di valore, come in

Der Sandmann di Hoffmann (1817). Di ragionamenti del genere sono piene sia la fantascienza, sia la saggistica contemporanea. Ciò che è interessante, nella maggior parte dei casi, è che in questo tipo di discorsi l'IA viene spesso rappresentata come una forma di alterità assoluta, che sembra somigliarci ma, in realtà, diventa qualcosa di talmente diverso da noi da non provare più nessuna forma di attaccamento alla nostra specie. Per questa ragione, in alcune di queste narrazioni, essa ci uccide, reputandoci poco funzionali alla sua sopravvivenza e, a volte, anche a quella del mondo in generale. In altre, come in *Her* (Jonze 2014), semplicemente ci abbandona, andando in cerca di un senso della sua esistenza che noi non la possiamo aiutare a trovare. Comunque, alla base di tutto questo c'è di nuovo, molto spesso, una critica al capitalismo, visto che nello stesso *Her* gli uomini si innamorano di un'intelligenza artificiale incorporea perché si sono abituati a vivere soli e nell'alienazione, in una società frenetica dove ognuno è una monade che persegue il proprio piacere individuale acquistando prodotti e servizi di ogni tipo, mentre in *Ex Machina* (Garland 2015) o nei due film di *Blade Runner* (Scott 1982; Villeneuve 2017) gli esseri umani vengono uccisi dai robot che hanno creato per realizzare profitti e per soddisfare il proprio desiderio di dominio sul creato.

In alcuni casi, però, come di nuovo in *Her*, la sindrome di Olimpia non conduce alla rovina le persone, ma è giustificata dal fatto che l'attaccamento alla macchina consente loro di migliorare, di crescere e di apprendere cose che, altrimenti, non avrebbero mai saputo, né su di sé, né sul mondo circostante, diventando pienamente Soggetti. È questa la sostanza del terzo modello culturale della figura 3, quello più transumanistico, nel quale si racconta di esseri umani che si fondono con l'intelligenza artificiale e vi caricano la propria mente, come nel film *Transcendence* (Pfister 2014), oppure si lasciano crescere da essa, sempre al fine di evolversi e di trasformarsi in una nuova specie, come in *I am mother* (Sputore 2019). La fantascienza e la saggistica contemporanea sono piene anche di questo tipo di discorsi, che a giudicare da inchieste come quella di O'Connell (2018), sono particolarmente in voga nella Silicon Valley e tra i grandi magnati del settore tecnologico, che sognano così, un po' come nella narrazione di Tegmark già riportata in queste pagine, di divenire immortali o di rendere immortale

il genere umano. Anche in questo caso, come nel modello culturale del Leviatano che vi sta accanto nella figura 3, l'intelligenza artificiale viene concepita come uno strumento salvifico, per la sua capacità di ragionare in maniera razionale e utilitaristica, andando oltre i "limiti" della nostra emotività e del pensiero simbolico-utopico, che possono rivelarsi disfunzionali, soprattutto quando si tratta di capire come sopravvivere in un mondo caratterizzato da una sempre maggiore scarsità delle risorse, che solitamente fa da sfondo a questo genere di narrazioni.

6. Ancora al centro dell'universo?

Come anticipato nel primo paragrafo di questo lavoro, i modelli culturali che utilizziamo per domandarci chi siamo noi al tempo dell'intelligenza artificiale hanno evidentemente a che vedere con il grande interrogativo che ha dato origine all'umanesimo. Infatti, ciò che siamo soliti chiederci, nei testi in cui parliamo del nostro rapporto con l'IA, è se siamo o se saremo ancora al centro dell'universo, dato che le tecnologie a cui stiamo dando origine potrebbero mettere in discussione questa convinzione.

A questo proposito, le narrazioni più catastrofiche dicono che saremo soppiantati da macchine più intelligenti e più performanti di noi, che nella migliore delle ipotesi ci utilizzeranno per riprodursi o ci lasceranno vivere nella più totale indifferenza, mentre loro prospereranno (Tegmark *op. cit.*, pp. 213-262). Questa prospettiva vale anche negli scenari meno fantascientifici, in cui l'intelligenza artificiale non viene vista come un soggetto autonomo, ma come uno strumento controllato da uomini potenti per i propri fini. In questi casi, saranno questi ultimi a dominare e al centro dell'universo si troveranno loro, mentre noi finiremo ai margini, "ingegnerizzati" dai loro mezzi informatici. Non si tratterà più, dunque, secondo una logica molto attestata nelle nostre culture fin dal medioevo, di avvicinarci alla "divinità" per mezzo della nostra tecnologia (Schatzberg *op. cit.*, pp. 29-41), ma di asservirci a nuovi "dei" dotati dell'IA.

Però, come abbiamo visto, accanto a questo tipo di discorsi, ne circolano altri più rassicuranti, nei quali invece l'umanesimo rimane saldo

all'orizzonte, proprio grazie alla stessa intelligenza artificiale e a tutti gli strumenti digitali di cui ci serviamo, che ci consentono di elevarci, affinando la nostra conoscenza del mondo. A questo proposito, la letteratura scientifica è piena di riferimenti allo “*human centred design*” della stessa intelligenza artificiale (Floridi e Cabitza *op. cit.*), oppure della tecnologia in generale (Friedman e Hendry 2019), che in fondo servono per ricordare che se gli strumenti che realizziamo non tengono conto di ciò che per noi è davvero importante, rischiamo di essere eterodiretti, in funzione di valori in cui non ci riconosciamo e che, per questo, potrebbero portarci a dire di non sentirci, per l'appunto, umani. Ma se, al contrario, immaginiamo dei mezzi tecnici che incorporino la nostra prospettiva, allora questi ultimi ci consentiranno di sviluppare al meglio la nostra natura, che fondamentalmente è quella di un “animale” che non è completo se non si appoggia sulla tecnologia (Leroi-Gourhan 1943; Ferraris *op. cit.*).

Entrambi questi tipi di discorsi sono di matrice narrativa e vengono costruiti mettendo l'uomo o la macchina nel ruolo del Destinante, vale a dire di chi indirizza il “destino” dell'uno e dell'altra. Inoltre, riflettano sul rapporto tra le logiche del Destinante stesso e la soggettività degli individui, umani o artificiali, a cui quest'ultimo si rivolge. Quando la macchina è solo un mezzo nelle mani dei suoi produttori e dei suoi proprietari, allora ci domandiamo se questi ultimi desiderino essere i Destinanti della gente comune che è chiamata a utilizzare la loro intelligenza artificiale, oppure se vogliono essere i Destinatari delle persone, perché questo cambia decisamente ciò che significa essere umani oggi, a seconda che ci si trovi in una posizione o nell'altra.

In effetti, questo genere di riflessioni fa quasi sempre i conti con il capitalismo e con il neo-liberismo, dato che nella nostra epoca si dibatte molto circa l'opportunità che i valori su cui questi ultimi si fondano siano quelli a cui devono attingere i nostri Destinanti per determinare il nostro futuro, oppure, più radicalmente, se dobbiamo accettare di essere i loro Destinatari (Fisher 2009; Han 2014). In questo senso, inevitabilmente, il dibattito sul sistema tecno-sociale che stiamo creando con le tecnologie digitali e con l'intelligenza artificiale non è altro che un sottoinsieme di quello più ampio, circa il mondo a cui stiamo dando origine e che ci determinerà per l'avvenire, come si può verificare

facilmente, se si sfogliano i libri più recenti sulla stessa IA, come per esempio *Atlas of AI* (Crawford 2021). Quasi sempre, i loro autori si domandano se vogliamo davvero perpetuare nelle nostre macchine “intelligenti” una mentalità economicistica, una razionalità prettamente utilitaristica, una capacità di calcolare costi e benefici nello sfruttamento delle risorse dell’ambiente, una potenza di calcolo fortemente energivora, per accentrare tutto questo potere nelle mani di poche persone, animate dal desiderio del proprio guadagno in concorrenza con le altre. Le risposte, come si è potuto notare in queste pagine, sono le più diverse, alcune anche positive, visto che, in fondo, questo modo di pensare ha prodotto alcune forme di progresso. Comunque, prevalgono le letture critiche, dato che a molti osservatori non sembra più che il capitalismo e il neo-liberismo mettano effettivamente l’uomo al centro dell’universo, quantomeno non tutti gli uomini.

Anche nei discorsi ambientalisti, del resto, il tema dell’umanesimo è ormai drammaticamente centrale, visto che l’idea di poter dominare la natura e di essere, così, padroni del nostro destino, in un mondo che si piega docilmente al nostro volere, è entrata radicalmente in crisi. Viviamo in un’epoca in cui il contrasto tra la nostra prospettiva individuale, giusta o sbagliata che sia, e quella di una serie di Anti-Destinati minacciosi che vi si oppongono, è divenuto il fulcro dei nostri discorsi a proposito di chi siamo e di chi vorremmo diventare. La posta in gioco, come è evidente, è molto alta, perché in base al modo in cui ci raccontiamo e ci racconteremo come sviluppare certe tecnologie, che notoriamente sono tra i mezzi più importanti attraverso cui regoliamo il nostro rapporto col creato e con gli altri, daremo origine a società molto differenti. Ma, soprattutto, determineremo la nostra identità di esseri umani per molti anni a venire, forse per sempre, a giudicare dallo stato in cui versa il nostro pianeta.

Riferimenti bibliografici

AA.VV. (2018), *L’Intelligenza artificiale al Servizio del Cittadino. Sfide e Opportunità*, Libro Bianco del Governo Italiano sull’Intelligenza artificiale, Agid (Agenzia per l’Italia Digitale), Roma.

- ASIMOV I. (1950), *I, Robot*, Harper Voyager, London.
- BOSTROM N. (2014), *Superintelligence. Paths, Dangers, Strategies*, Oxford University Press, Oxford.
- CASILLI A. (2019), *En attendant les robots: enquête sur le travail du clic*, Seuil, Paris.
- CRAWFORD K. (2021), *Atlas of AI*, Yale University Press, London.
- FERRARIS M. (2021), *Documanità. Filosofia del mondo nuovo*, Laterza, Roma.
- FERRARO G. (2001), *Il linguaggio del mito*, Meltemi, Roma.
- FISHER M. (2009), *Capitalist Realism: Is There No Alternative?*, Zero Books, London.
- FLOCH J.M. (1990), *Sémiotique, marketing et communication. Sous les signes, les stratégies*, Presses Universitaires de France, Paris.
- FLORIDI L., F. CABITZA (2021), *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Milano.
- FRIEDMAN B., D.G. HENDRY (2019), *Value Sensitive Design. Shaping Technology with Moral Imagination*, Mit Press, London.
- FRISCHMANN B., E. SELINGER (2019), *Re-Engineering Humanity*, Cambridge University Press, Cambridge.
- FRY H. (2018), *Hello World. Being Human in the Age of the Machine*, Black Swan, London.
- GIOVANNOLI, R. (2015), *La scienza della fantascienza*, Bompiani, Milano.
- GREIMAS A.J. (1970) *Du sens*, Éditions du Seuil, Paris.
- GREIMAS A.J., J. FONTANILLE (1991), *Sémiotique des passions. Des états des choses aux états d'âme*, Éditions du Seuil, Paris.
- HAN B.C. (2014), *Psychopolitic: Neoliberalismus und die neuen Machttechniken*, Fischer Velrlaag, Frankfurt am Main.
- HARARI Y. (2015), *Homo Deus. A Brief History of Tomorrow*, Penguin, London.
- HOFFMANN E.T. A. (1817), *Der Sandmann*, Callots Manier, Berlin.
- KAPLAN J. (2016), *Artificial Intelligence. What Everyone Needs to Know*, Oxford University Press, London.
- KURZVEIL R. (2005), *The Singularity is Near. When Humans Transcend Biology*, Duckworth Overlook, London.
- LANDOWSKI E. (1997), *La prospettiva socio-semiotica*, in Lexia n.6: 5-7.
- LEROI-GOURHAN, A (1943), *L'homme et la matière. Évolution et techniques*, I, Albin Michel, Paris.
- NEGROPONTE N. (1995), *Being Digital*, Vintage, New York.

- O'CONNELL M. (2018), *To Be a Machine. Adventures Among Cyborgs, Utopians, Hackers and the Futurists Solving the Modest Problem of Death*, Granta Books, London.
- O'NEIL K. (2016), *Weapons of math destruction. How big data increases inequality and threatens democracy*, Penguin, New York.
- PASQUALE F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge.
- (2020), *New Laws of Robotics. Defending Human Expertise in the Age of AI*, The Belknap Press of Harvard University, Cambridge.
- PEZZINI I. (a cura di) (1991), *Semiotica delle passioni*, Esculapio, Bologna
- PROPP V. (1928), *Morfologija skazki*, Academia, St. Petersburg.
- SADIN E. (2018), *L'intelligence artificielle ou l'enjeu du siècle: Anatomie d'un antihumanisme radical*, L'Echappée, Paris (trad. it. *Critica della ragione artificiale. Una difesa dell'umanità*, Luiss, Roma, 2019).
- SANTANGELO A. (2013), *Sociosemiotica dell'audiovisivo*, Aracne, Roma, 2013
- (2021 a), *Volti simbolici. Per una teoria sociosemiotica del volto*, in *Lexia* 37/38: 503–520.
- (2021 b), *Culturally significant symbolic faces. For a sociosemiotics of the face*, in *Signs, systems, studies*, 49 (3/4): 418–436.
- SCHATZBERG E. (2018), *Technology. Critical History of a Concept*, The University of Chicago Press, London, 2018.
- SHELLEY M. (1818), *Frankenstein. Or the modern Prometheus*, Lackington, London.
- SRNICECK N., A. WILLIAMS (2015), *Inventing the Future: Postcapitalism and a World without Work*, Verso Books, London.
- TEGMARK M. (2017), *Life 3.0. Being Human in the Age of AI*, Penguin, London.
- ZUBOFF S. (2019), *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, Profile Books, London.

INTERPRETARE IL CAMBIAMENTO CLIMATICO ATTRAVERSO L'INTELLIGENZA ARTIFICIALE UNA PROSPETTIVA SEMIOTICA

CRISTINA VOTO*

No hay que ver el futuro para saber lo que va a pasar

No (Sirens) Nicolas Jaar, 2016

ENGLISH TITLE: *Interpreting climate change through artificial intelligence. A semiotic perspective change.*

ABSTRACT: These pages aim to reflect on the defining characteristics of artificial intelligence in order to account for the interpretive habits with which we relate to scientific knowledge when it passes through systems using these technologies. From this initial position, we set up a semiotic gaze to account for the challenges we face using artificial intelligence systems for climate change risk assessment. Thinking of big data collections on climate change as archives that, depending on selection criteria, can become the databases needed to train artificial intelligence, will bring out the need to reflect on the epistemic frontier that organizes data.

KEYWORDS: Archive, Climate Change, Database, Data Visualization, Uncertainty

1. Introduzione. Di cosa parliamo quando parliamo di intelligenza artificiale?

Le intelligenze artificiali (IA) sono quei sistemi tecnologici che, in vista del raggiungimento di obiettivi specifici, possono compiere azioni con un certo grado di autonomia per mezzo di un'analisi computazionale. Sono sistemi complessi, oggi più che mai pervasivi, capaci di plasmare discorsi culturali, scientifici, produttivi ed etici e, allo stesso tempo, di innervare le ontologie che ne modellano i costrutti. Tuttavia,

* Università degli Studi di Torino.

recuperando una celebre espressione di Kate Crawford, questi sistemi non possono essere definiti né intelligenti e men che meno artificiali (2021). Come scrive l'autrice, l'IA: "nasce dai laghi salati della Bolivia e dalle miniere del Congo, ed è costruita a partire da set di dati etichettati da crowdworkers che cercano di classificare azioni, emozioni e identità umane." (Crawford *op. cit.*, p. 250). Nasce, cioè, da un processo di sfruttamento estrattivo delle risorse naturali e dalla precarizzazione del lavoro cognitivo (Berardi 1995) che sostiene la cosiddetta *gig economy*⁽¹⁾. In questo senso, nonostante il tentativo di astrarne una lettura ideale e oggettiva attraverso immaginari tecnocratici di automazione ed efficienza, per capire il fenomeno dell'IA è necessario restituirne la dimensione opaca che organizza le infrastrutture connettive e regola le dinamiche biopolitiche (Voto 2021). Inoltre se da un punto di vista storico, e secondo un *topos* oramai consolidato, il documento firmato a otto mani nel 1955 da J. McCarthy (Dartmouth College), M.L. Minsky (Harvard University), N. Rochester (I.B.M. Corporation), e C.E. Shannon (Bell Telephone Laboratories) con cui si presenta la *Proposta per il progetto estivo di ricerca sull'intelligenza artificiale di Dartmouth* può considerarsi il testo inaugurativo degli studi che portarono allo sviluppo delle IA così come la conosciamo oggi (Chun 2021; Crawford *op. cit.*; Larson 2021; Floridi 2022), sin dai suoi albori il fenomeno si caratterizza per la sua natura "controfattuale" rispetto alla qualità intelligente (Floridi *op. cit.*, p. 44). Come possiamo leggere nell'introduzione al testo, per i quattro autori lo scopo della ricerca è: "far sì che una macchina agisca con modalità che sarebbero definite intelligenti se un essere umano si comportasse allo stesso modo" (McCarthy, Minsky, Rochester, Shannon [1955] 2006, p. 11)⁽²⁾. L'obiettivo perseguito dai firmatari dello studio non è la riflessione sulle possibili caratteristiche intelligenti delle macchine e/o sul loro eventuale sviluppo. Diversamente da quelle condotte da Alan Turing (1950), le ricerche di McCarthy, Minsky,

(1) L'enciclopedia online Treccani si definisce la *gig economy* come il "modello economico basato sul lavoro a chiamata, occasionale e temporaneo, e non sulle prestazioni lavorative stabili e continuative, caratterizzate da maggiori garanzie contrattuali" (https://www.treccani.it/vocabolario/gig-economy_%28Neologismi%29/, ultimo accesso 23/07/2022).

(2) Nell'originale: "For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving".

Rochester e Shannon danno avvio allo sviluppo di una forma dell'agire che può: “affrontare con successo compiti e problemi, in vista di un obiettivo, senza alcun bisogno di essere intelligente” (Floridi *op. cit.*, p. 91). Il successo di questa *mindless agency*, come la definisce Luciano Floridi, nella risoluzione dei compiti assegnati è agevolato dal progressivo rimodellamento dei contesti per mezzo di un avvolgimento (*ibid.*) degli ambiti in cui l'IA è operativa. La divergenza tra azione e intelligenza associata al rimodellamento delle ontologie ci restituisce “un approccio al mondo che non è né *descrittivo* né *prescrittivo*” (*ibid.*, p. 54, *italica nell'originale*) ma piuttosto vincolante perché in grado di ri-ontologizzare il mondo circostante.

Queste prime osservazioni ci servono per collocare le pagine che seguono entro un orizzonte epistemico preciso. Se l'IA non è qualcosa che si possa facilmente definire ma soprattutto se le caratteristiche definitorie che la denotano ad oggi, ovvero l'artificialità e l'intelligenza, si prestano ad interpretazioni equivoche, crediamo che un punto di partenza per la comprensione del fenomeno possa essere quello di ripartire dalle pratiche. Se prendiamo per buona l'ipotesi per cui l'IA *fa cose*, ingaggiare la capacità semiotica di cogliere il senso nel suo dispiegarsi può aiutarci ad analizzare come l'IA agisce sul mondo. Per questa ragione, l'obiettivo delle pagine che seguono sarà quello di riflettere su un agire specifico dell'IA e, nello specifico dell'apprendimento automatizzato ovvero del *Machine Learning*: la valutazione visiva del rischio riguardo il cambiamento climatico. Del resto, oggi “non esiste in pratica alcuna analisi scientifica o politica basata su evidenze che non sia alimentata da tecnologie digitali avanzate. In questo quadro, l'IA sta diventando parte integrante degli strumenti necessari per far progredire la nostra comprensione scientifica in molti ambiti” (Floridi 2021: 299).

2. *Habits* e modelli per pensare il cambiamento climatico

Con il sintagma “cambiamento climatico” ci si riferisce alle variazioni a lungo termine delle temperature e dei modelli metereologici⁽³⁾. Sebbene

(3) Riportiamo la definizione che ne dà il sito web delle Nazioni Unite <https://unric.org/it/che-cosa-sono-i-cambiamenti-climatici/>, ultimo accesso il 23/07/2022.

questi mutamenti possano avere anche cause naturali — è la Storia a raccontarci di piccole e grandi ere glaciali come quella successa durante l'Olocene o tra il XIV e il XIX secolo —, nel caso del cambiamento climatico in seno all'attuale riscaldamento globale risulta difficile non decretare la responsabilità umana come fattore condizionante e incidente⁽⁴⁾. È stato Frank Luntz, consulente dell'allora presidente statunitense George W. Bush, a suggerire l'introduzione dell'espressione *cambiamento climatico* al posto di *riscaldamento globale* (Luntz 2002), un'operazione di successo retorico che partiva dal constatare che “la descrizione dei fatti è così pericolosamente prossima alla prescrizione di una politica che, per arrestare la messa in discussione dello stile di vita industriale, sono i fatti a dover essere messi in dubbio” (Latour [2015] 2020, p. 34).

Da un punto di vista enunciativo, asserire l'esistenza di un'emergenza di causa antropica comporta una valutazione del rischio tale per cui la virtualità dei modelli predittivi usati per la comunicazione dello stesso può essere in grado di prescrivere una condotta preventiva. Proprio in ragione di questa carica modale, la valutazione del rischio è una stima difficilmente oggettiva quanto, piuttosto, il risultato di un posizionamento situato culturalmente. Valutare il rischio comporta, in fin dei conti, legittimare un cambiamento negli *habits* interpretativi⁽⁵⁾, ovvero accettare che una disposizione condivisa e generalizzata, acquisita attraverso catene di interpretazioni associative o dissociative, debba essere corretta affinché possa produrre abiti d'azione, nel nostro caso, più sostenibili. Se prendiamo per buona questa prospettiva, la riflessione sugli *habits* può rilevarsi uno strumento utile al contrasto del dilagante negazionismo climatico, un discorso questo che intravede nell'attenzione diretta sull'emergenza ambientale una cospirazione a carico di scienziati e governi per il controllo delle politiche globali. Questa disposizione, la cui validità discorsiva risiederebbe nel riconoscimento dell'impossibilità, da parte della comunità scientifica e delle *governance*, di progettare

(4) Dobbiamo al fisico francese Jean-Baptiste Joseph de Fourier (1768–1830) la prima formulazione scientifica riguardo i cambiamenti climatici: nelle sue ricerche egli iniziò a paragonare l'atmosfera terrestre a una serra per via della sua capacità di trattenere il calore. È invece ad opera di Gilbert Plass (1920–2004) il primo studio sul legame tra anidride carbonica e cambiamento climatico.

(5) Il riferimento è alla semiotica di Charles S. Peirce.

un rilevamento oggettivo del rischio, può essere modificata attraverso la legittimazione della valutazione come strumento interpretativo e situato delle conseguenze del cambiamento climatico.

Un caso esemplificativo per riflettere sul limite interpretativo del negazionismo è lo scandalo scoppiato intorno al cosiddetto *Climategate* del 2009. La vicenda ha in realtà inizio nel 1998 quando i paleoclimatologi Michael E. Mann, Raymond S. Bradley e Malcom K. Hughes pubblicano il grafico di tendenze conosciuto come *Hockey Stick Chart* (1998). Si tratta una simulazione dei cambiamenti della temperatura media nell'emisfero settentrionale tra il 1400 e il 1995 aggiornata l'anno successivo al periodo compreso tra il 1000 e il 1998 (1999). Se da un punto di vista semiotico, e nello specifico appellando alla riflessione sulla natura iconica del segno portata avanti da Charles S. Peirce, possiamo dire che ogni grafico, in qualità di diagramma, si configura sulla base della relazionalità con cui si interfacciano i processi cognitivi rappresentati, la lettura dell'*Hockey Stick Chart* che ne fa il discorso negazionista riconosce nel grafico una traduzione fenomenica dell'oggetto che rappresenta.

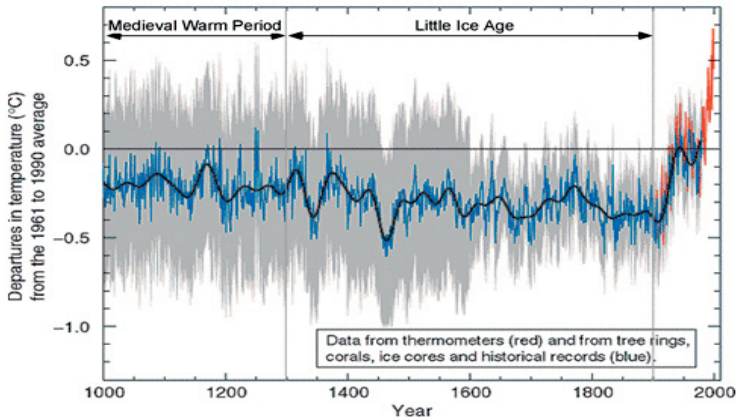


Figure 1. Mann's "hockey stick" temperature history for the past 1,000 years, including region of uncertainty (gray area) and time periods of the Medieval Warm Period and Little Ice Age—events that are largely absent in this record. (adapted from the IPCC's Third Assessment Report).

Figura 1. Il grafico conosciuto come Hockey Stick Chart (1999).

Nel 1999 il grafico venne inserito nel primo rapporto firmato dall'*Intergovernmental Panel on Climatic Change* (IPCC)⁽⁶⁾ e nella sezione *Summary for Policy Makers* per essere poi utilizzato dall'allora presidente degli Stati Uniti Bill Clinton e dal suo vicepresidente Al Gore come prova dell'impatto antropico sul clima. A partire da allora l'*Hockey Stick* fu attaccato da varie schiere politiche così come da scienziati fisici e sociali sulle pagine di prestigiose riviste quali il *Wall Street Journal* (dove il 14 febbraio 2005 si pubblica l'articolo "In Climate Debate, The 'Hockey Stick' Leads to a Face-Off Nonscientist Assails a Graph Environmentalists Use, and He Gets a Hearing") e il *MIT Technology Review* (con la pubblicazione dell'articolo "The Hockey Stick" l'11 maggio del 2005) in ragione di due argomenti: uno a sfondo più prettamente cognitivo, l'altro epistemico.

Al dare forma alle tendenze del clima nel periodo compreso tra il 1000 il 1400, il grafico visualizza i dati delle temperature durante il cosiddetto *riscaldamento medievale* un periodo, cioè, di alterazione dei livelli climatici in cui l'impatto antropico sull'ambiente è inesistente. Gli attacchi al grafico si concentrarono perciò sulla validità d'uso dei metodi statistici impiegati nella ricostruzione della temperatura media, nello specifico l'analisi delle componenti principali⁽⁷⁾ (Chun 2015), e sulla selezione dei parametri (Walsh 2014). I dati presentati nel grafico *Hockey Stick* sono il risultato di parametrizzazioni indirette ottenute dagli *archivi della Terra*, quell'insieme di informazioni ricavabili dalla misurazione degli anelli degli alberi, dal carotaggio del ghiaccio e dai co-

(6) Nel 1988 l'Organizzazione Meteorologica Mondiale (WMO) e il Programma delle Nazioni Unite per l'Ambiente (UNEP) fondano l'Intergovernmental Panel on Climatic Change (IPCC), un foro scientifico che oggi riunisce 195 stati membri e che si prefigge come obiettivo principale quello di divulgare valutazioni periodiche riguardo le cause e i rischi del cambiamento climatico. Attraverso report periodici (meglio conosciuti come *Assesment Report* riassunti poi nei *Summary for Policy Makers*) l'IPCC progetta e diffonde materiali sulle proiezioni a futuro dei mutamenti attraverso grafici che simulano opzioni di risposta sulla base di diversi modelli di visualizzazione di dati. Il primo report è del 1990 e una costante di queste testualità è quella di accompagnare le relazioni scritte con infografiche quali grafici a linee, linee di tendenza e scenari.

(7) L'analisi delle componenti principali è una tecnica statistica sviluppata all'inizio del ventesimo secolo da Karl Pearson, padre della moderna statistica moderna, biometrico ed eugenetico. L'Analisi delle Componenti Principali risolve un insieme di punti di dati possibilmente correlati in un insieme di componenti principali determinando gli autovettori della matrice di correlazione. Attraverso una scomposizione dei vettori l'analisi delle componenti Principali riorganizza i dati attorno a un nuovo insieme di assi, il che rende più facili i calcoli.

ralli; successivamente messi a confronto con le informazioni ricavabili dalla registrazione delle anomalie climatiche iscritte negli *archivi della Società*, come annali, documenti o testimonianze, etc. La critica mossa agli autori del grafico è, allora, quella di aver attinto a diversi tipi di parametri per la modellizzazione della temperatura secondo un campionamento carente in uniformità temporale e spaziale. In particolare i dati ricavati dalle variazioni degli anelli degli alberi furono messi sotto torchio: questi parametri descrivono infatti dati riconoscibili solo nei continenti di media latitudine ragione per cui ampie zone del globo come i poli o gli oceani non possono essere adeguatamente rappresentate. In ogni caso, però, le linee del grafico ci restituiscono una tendenza precisa: quando le tecniche di rilevamento dei dati diventano più efficaci le temperature dell'emisfero boreale si innalzano, da cui il riferimento eidetico alla mazza da hockey, dove l'impennata delle temperature diventa la lama della mazza.

Seppur si possa affermare che “l'accuratezza nella restituzione della complessità in questo grafico sia stata sacrificata sull'altare dell'urgenza di comunicare un concetto molto chiaro e imperativo: le temperature si stanno impennando e non c'è più tempo da perdere” (Burgio 2021, p. 129–130), il caso del *Climategate* ci mostra che il discorso negazionista si radica in quell'*habit* interpretativo che fa della scienza un sapere disincarnato (Latour 1987). Contrastare questa lettura comporta, perciò, riconoscere la validità di una conoscenza scientifica che possa — anzi, debba — costruirsi progressivamente a partire dal lavoro di controllo sulla traduzione, scrittura e diffusione dei dati compiuto da quelle istituzioni dove “i fatti sono insieme accuratamente fabbricati e resi fattuali grazie alla cura riservata a una simile fabbricazione” (Latour [2015] 2020, p. 179). Inoltre, permette il riconoscimento di un'ulteriore disposizione condivisa che fa del discorso scientifico un sigillo di veridizione piuttosto che un agente di affidabilità. Dal nostro punto di vista si tratterebbe, invece, di passare da un'idea di validità scientifica come aderenza al reale per far posto al riconoscimento di una validità basata sulla solidità delle ipotesi.

Le vicissitudini legate alla ricezione dell'*Hockey Stick Chart* possono servirci, allora, da punto di partenza per riflettere sui tipi di conoscenza che le attuali convenzioni di visualizzazione possono esprimere.

Esiste un atteggiamento diffuso che fa della visualizzazione dei dati un processo trasparente (Walsh 2014) tale per cui dati e visualizzazione risulterebbero la stessa cosa. Tuttavia un'espressione grafica può essere utilizzata non solo per visualizzare ma anche per creare o mostrare le caratteristiche che fanno di una determinata forma diagrammatica un modello di attività interpretativa. L'accento sulla modellazione mette in dubbio l'unidirezionalità tra dati e visualizzazione e ne trasforma il processo in uno scambio bidirezionale (Drucker 2020). Ed è forse questa la lacuna principale cui ci mette a confronto il grafico a mazza da hockey: l'assenza di una riflessione sulla reciprocità strutturale e materiale tra grafico e dati. Sostenere, invece, la necessità di una interpretazione dei modelli, così come proposta da Johanna Drucker, comporta accettare che un'espressione grafica possa servire come modalità primaria di produzione della conoscenza e non come espressione secondaria di dati preesistenti. In questo senso: "La modellazione dell'interpretazione richiede forme grafiche diverse e distinte, capaci di esprimere ambiguità, contraddizioni, sfumature, cambiamenti e altri aspetti della considerazione critica. Ciò richiede forme e formati non conformi agli standard delle metriche regolari su griglie di coordinate governate da un sistema cartesiano standard" (Drucker *op. cit.*, pp. 99–100, traduzione nostra). Riflettere su come i dati dovrebbero apparire comporta passare da un'idea di visualizzazione a quella di modellazione dell'interpretazione verso il raggiungimento di una *graphicacy*, ovvero un'alfabetizzazione grafica che possa alimentare la capacità di comprendere, utilizzare o generare immagini grafiche, tale da rendere conto della natura complessa e situata del fenomeno in analisi.

3. Dagli archivi ai database: i dati come strumento per interpretare l'incertezza

I dati attraverso cui elaboriamo la valutazione del rischio del cambiamento climatico sono estratti da un apparato di rilevamento e telerilevamento globale (Bratton 2019), composto da una fitta maglia di sensori, più o meno remoti, necessari alla raccolta di informazioni. In questo senso, e come ci ricorda Wolfgang Behringer nel suo *Storia culturale del clima*:

con la nascita dell'elaborazione elettronica dei dati non solo nacquero le previsioni del tempo, ma si cominciò anche a formulare pronostici sull'evoluzione del clima. Dopo il primo atterraggio lunare, nel 1969, il satellite Nimbus III permise di misurare la temperatura in tutto il mondo. Ma solo nel 1970 si misero a punto i primi modelli climatici complessi a più variabili. (Behringer [2010] 2013, p. 370)

Oggi esiste un'enorme quantità di dati geocodificati capaci di informarci sulle interazioni tra essere umano e ambiente da cui ottenere informazioni preziose per il monitoraggio della crescita urbana, dei cambiamenti nei modelli di utilizzo del suolo, dell'erosione costiera, della deforestazione, della torbidità dell'acqua, nonché per la stima di parametri fisici come la temperatura degli oceani e/o l'identificazione di oggetti sulla superficie terrestre. Inoltre, le immagini di telerilevamento ad alta risoluzione di satelliti e droni possono consentire una mappatura dinamica della popolazione che può essere, a sua volta, utilizzata per valutare il rischio di disastri naturali, tracciando le tendenze dei modelli di esposizione nel tempo. E a questi dati ogni giorno se ne aggiungono altri, compresi i big data provenienti dai post dei social media e dalle registrazioni dei dati dei telefoni cellulari.

Questi grandi archivi di informazioni, una volta organizzati nelle banche dati, sono usati per alimentare sistemi di *Machine Learning*, spesso con un obiettivo preciso: affrontare l'incertezza derivante dalla non linearità del sistema atmosferico e dalla natura complessa delle variabili climatiche. L'utilizzo del *Machine Learning* nella valutazione del rischio del cambiamento climatico è con sempre maggior frequenza legato a compiti di modellazione e di previsione delle vulnerabilità legati alla dimensione antropica, espressi nei termini di condizioni demografiche, vulnerabilità della dimensione ambientale a causa di inondazioni, alluvioni improvvise, frane, mareggiate ed eventi ventosi.

Riflettere sull'incidenza antropica nei confronti del clima così come progettare modelli interpretativi del mutamento climatico ci mette, perciò, a confronto con una delle più grandi sfide della contemporaneità. Sono sfide prometeiche sia da un punto di vista scalare, perché ad essere in gioco è la sussistenza del Pianeta così come lo conosciamo, sia dal punto di vista delle relazioni geopolitiche, perché ad essere

necessario è il consenso internazionale sulle misure da adottare. Il riferimento a Prometeo non è casuale: in una conferenza datata 2008, Bruno Latour propone di ripensare la società digitale e la riprogettazione che comporta la smaterializzazione del significante proprio a partire da questa figura mitologica. Oggi quando “ogni singola cosa, ogni dettaglio della nostra esistenza quotidiana, dal modo in cui produciamo cibo, al modo in cui viaggiamo, al modo in cui costruiamo automobili o case, al modo in cui cloniamo mucche, ecc. deve essere riprogettato” (Latour [2008] 2009, p. 256), ripartire dal titano figlio di Giapeto e di Climene permette di inquadrare il ruolo dell’“esplicitazione” come strumento per la “rimaterializzazione” (*op. cit.*). L’esplicitazione titanica è qui da intendersi come risposta all’avviluppamento che viviamo in una società dove “non siamo mai fuori senza aver ricreato un altro involucro più artificiale, più delicato, più elaborato” (Latour *op. cit.*, p.259).

Anche queste pagine nascono dall’inquietudine verso l’esplicitazione dei processi di ri-ontologizzazione attraverso cui stiamo trasformando il mondo così come lo conosciamo per essere sempre più adatto al funzionamento dell’IA. Nello specifico, è la domanda verso lo statuto pragmatico dei database, ovvero i grandi archivi di dati, ad essere al centro della nostra riflessione.

Dal nostro punto di vista, per quanto riguarda il problema del cambiamento climatico, bisognerà esplicitare l’avviluppamento che riguarda almeno tre forme-archivio capaci di diventare banche di dati per il rimodellamento ontologico degli enunciati:

- gli *archivi della Terra*, ovvero le informazioni sul clima che possono essere ricavate sia dai telerilevamenti sia dai sedimenti naturali come quelli presenti nei coralli, nelle venature degli anelli di accrescimento degli alberi o nel carotaggio del ghiaccio, etc.;
- gli *archivi della società*, i registri delle trasmissioni consapevoli di informazioni riguardo i fenomeni climatici;
- gli *archivi del clima*, le grandi raccolte di dati con cui immaginare i modelli grazie a sistemi di telerilevamento come droni o satelliti.

Questi insiemi strutturati di dati decretano sia la storicità sia la trasformabilità degli enunciati che possiamo immaginare riguardo il

cambiamento climatico. Con la proposta di utilizzare la categoria di archivio per pensare l'articolazione tra sistemi di IA e valutazione del rischio del riscaldamento globale, stiamo recuperando la posizione teorica di Patrizia Violi (2014), per cui ogni archivio può essere studiato sia come *iconosfera*, un'enciclopedia visiva legata a un determinato evento; sia, nei termini di Michel Foucault (1969), come un dispositivo di dicibilità capace di generare nuove forme discorsive e pratiche socioculturali. Ripartire da Violi ci serve, inoltre, per fare luce su un aspetto fondamentale della semiotica peirceana: per il filosofo statunitense anche le espressioni matematiche sono icone⁽⁸⁾. Nelle espressioni matematiche è infatti possibile individuare una relazione diagrammatica con l'oggetto che permette la scoperta di: "verità nuove oltre a quelle che sono sufficienti a determinare la costruzione dell'icona stessa" (Peirce, C.P. 2.279). Ecco che allora che anche i dati di rivelamento strumentali collezionati nelle forme-archivio possono essere analizzati come elementi appartenenti all'iconosfere utili per la conoscenza di: "nuovi aspetti dell'Oggetto, proprio attraverso l'osservazione e la manipolazione delle relazioni" (Proni 2007, p. 8). Nelle iconosfere che conformano gli archivi della Terra, della società e del clima troveremo, allora, non solo grafici e immagini ma anche i dati che, in quanto espressioni matematiche, sono necessari alla costruzione di forme discorsive del cambiamento climatico.

Ma come funzionano questi archivi? La definizione di segno fornita da Charles S. Peirce (C.P. 2.218), può tornarci utile: un archivio, in quanto segno, è qualcosa che sta a qualcuno — un certo criterio di selezione come per esempio, nel nostro caso, la parametrizzazione attraverso cui è stato svolto il rilievo strumentale dei dati — per qualcosa — ovvero un dato concreto o una collezione di dati concreti che operino come sostituti del fenomeno in analisi — in qualche aspetto o capacità — secondo, cioè, una certa classificazione. Costruire un sistema di IA, cioè, ci mette sempre a confronto con le frontiere epistemiche che regolano il funzionamento dei set di dati: il limite tra ciò che sarà archiviato e ciò che non lo sarà sulla base dei criteri di classificazione.

In precedenti studi sulla performatività degli archivi (Acebal, Guerri,

(8) Come afferma Peirce: "Thus, an algebraic formula is an icon, rendered such by the rules of commutation, association, and distribution of the symbols" (C.P. 2.279).

Voto 2020), applicando la Teoria degli atti linguistici (Austin 1962, Searle 1969) e recuperando il concetto austiniano di *fallacia descrittiva del linguaggio*, abbiamo sostenuto la *fallacia informativa dell'archivio*. L'archivio, una volta che diventa il database con cui alimentare un'IA, produce conoscenza e rimodella il problema da risolvere attraverso gli avviluppamenti che organizzano la struttura dei dati. Inoltre, la posta in gioco epistemica aumenta drammaticamente, così come le ripercussioni etiche nella progettazione dell'interazione essere umano-computer, quando l'IA non è programmata con correlazioni logiche, ma è autorizzata a programmare autonomamente, come nel caso del *Machine Learning*. Nell'ambito di queste tipologie di IA, l'esperienza di apprendimento viene addestrata su insiemi di dati che fungono da conoscenza empirica archiviata e a cui il sistema ha accesso. L'addestramento consiste, infatti, nel parametrizzare una funzione per i segnali di input in modo da produrre l'informazione attesa in uscita e, in caso di risultati inattesi, sarà la tecnologia a dover regolare autonomamente i propri output. I dataset sono quindi fondamentali per acquisire la competenza e le prestazioni necessarie al *Machine Learning*, poiché il sistema non può acquisire conoscenza al di là dei dati con cui è alimentato, a meno che non provenga *a priori*, codificata nei dati stessi o nel sistema.

Affrontare il problema del cambiamento climatico significa, allora, non solo riflettere sulle conseguenze ecologiche dell'Antropocene, quanto piuttosto mettere alla prova i modelli epistemici esistenti per la lettura e interpretazione dei dati con la finalità di superare le polarizzazioni degli *habits* del negazionismo (che funzionano sia sulla base di un'opposizione natura VS cultura, sia di un'opposizione tra scienza disincarnata VS scienza incarnata) verso una prospettiva *more-than-human*. Significa cioè poter ricorrere a una prospettiva post-umana in cui l'IA è alleato fondamentale per la modellazione interpretativa dell'emergenza climatica⁽⁹⁾.

(9) Seppure non saranno queste le pagine dove affronteremo il problema dell'emissioni di gas serra provocati dai processi di addestramento delle IA siamo coscienti della gravità di questo aspetto del fenomeno in analisi, rimandiamo ancora una volta a Floridi (2022) per un maggiore approfondimento sull'argomento.

4. Conclusioni

Nell'arco di queste pagine abbiamo tracciato un percorso che da una messa in discussione delle qualità definitorie dell'IA, ovvero l'artificialità e l'intelligenza, ci ha portate verso un ripensamento degli *habits* interpretativi con cui ci relazioniamo nei confronti della conoscenza scientifica. In questo percorso, poi, abbiamo gettato le basi per l'organizzazione di una struttura analitica con cui affrontare le sfide dinnanzi a cui ci mettono di fronte i rischi del cambiamento climatico. Pensare alle grandi collezioni di dati come archivi che, in base ai criteri di selezione, possono diventare i database necessari per l'addestramento delle IA ha evidenziato la necessità di riflettere sulla frontiera epistemica che organizza questi insiemi. Arrivate a questo punto resta, forse, solo un'ultima riflessione, che può dare spazio a ulteriori e nuove ricerche. Le simulazioni, come quelle compiute dai sistemi di *Machine Learning* per la valutazione del rischio di un'emergenza planetaria come il cambiamento climatico, sono altamente imprevedibili. Quest'ultima osservazione, allora, ci permette di mettere a fuoco il ruolo dell'analisi computazionale nel decifrare la struttura degli spazi di possibilità (De Landa 2002), gli interstizi virtuali che esistono tra i dati. Compito delle ricerche future, allora, sarà quello di mettere in funzione un meccanismo che permetta l'emergenza di questi spazi dall'incertezza, perché incerti in realtà non lo sono affatto.

Riferimenti bibliografici

- ACEBAL M. e C. GUERRI e C. VOTO (2020), "The Performativity of the Archive from a Semiotic Perspective", *Southern Semiotic Review*, 13: 32 — 47 http://dx.doi.org/10.33234/SSR/13_3_2020.
- AUSTIN, J. (1962), *How to Do Things with Words*, Oxford University Press, Oxford.
- BEHRINGER, W. (2010), *Kulturgeschichte des Klimas. Von der Eiszeit zur globalen Erwärmung*, Munich, Verlag (trad. it *Storia culturale del clima. dall'Era glaciale al Riscaldamento globale*, Torino, Bollati Boringhieri).
- BERARDI, F. (1995), *Neuromagma. Lavoro cognitivo e infoproduzione*, Castelvecchi, Roma.

- CAIRO A. (2016), *The Truthful Art. Data, Charts, and Maps for Communication*.
- CHUN W. (2021), *Discriminating Data. Correlation, Neighborhoods, and the New Politics of Recognition*, Cambridge MA, MIT Press.
- CHUN W. (2015), “On Hypo–Real Models, or Global Climate Change: A Challenge for the Humanities”. *Critical Inquiry*, 41(3): 675 — 703 <https://doi.org/10.1086/680090>.
- CRAWFORD K. (2021), *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven and London, Yale University Press (trad. It. *Né intelligente né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino 2021).
- DE LANDA M. (2002), *Philosophy and Simulation: The Emergence of Synthetic Reason*, London, Bloomsbury.
- MCCARTHY J., M.L. Minsky, N. Rochester, C.E. Shannon (1955), “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, *AI Magazine*, 27(4): 1 — 12 <https://doi.org/10.1609/aimag.v27i4.1904>.
- DRUCKER J. (2020), *Visualization and Interpretation. Humanistic Approaches to Display*, Cambridge MA, MIT Press.
- FLORIDI L. (2022), *Etica dell'intelligenza artificiale: sviluppi, opportunità, sfide*, Milano Raffaele Cortina.
- FOUCAULT M. (1969), *L'Archéologie du savoir*, Paris, Gallimard.
- LARSON E.J. (2021), *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, Harvard, Harvard University Press.
- LATOUR B. (1987), *Science in Action: How to Follow Scientists and Engineers Through Society*, Harvard, Harvard University Press.
- (2008), “A Cautious Prometheus? A Few Steps Toward a Philosophy of Design with Special Attention to Peter Sloterdijk”. In W. SCHINKEL, L. NOORDEGRAAF–EELENS, L. TSIPOURI, V. STENIUS (Eds.), *In Medias Res: Peter Sloterdijk's Spherological Poetics of Being*, Amsterdam, Amsterdam University Press (trad. it. “Un Prometeo cauto? Primi passi verso una filosofia del design”, *E/C: rivista dell'Associazione Italiana di Studi Semiotici*, 3(4): 255 — 263).
- (2015), *Face à Gaïa. Huit conférences sur le nouveau régime climatique*, Paris, La Découverte (trad. it. *La sfida di Gaia. Il nuovo regime climatico*, Milano, Meltemi 2020).
- LUNTZ F. (2002), “The Environment: A Cleaner, Safer, Healthier America”, Luntz Research Companies <https://www.the-republican-reversal.com/uploads/1/2/0/2/120201024/luntzresearch.memo2.pdf>

- MANN M.E., R.S. BRADLEY, M.H. HUGHES (1998), "Global-scale temperature patterns and climate forcing over the past six centuries", *Nature*, 392: 779–787.
- (1999), "Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations", *Geophysical Research Letters* 26(6): 759–762.
- PEIRCE C.S. (1931–1935), *Collected Papers of Charles Sanders Peirce*, 8 vol., ed. Hartshore, Weiss et Burks, Cambridge MA, Harvard University Press
- PRONI G. (2007), "L'icona nella teoria di Peirce come condizione della conoscenza e del segno". In G. PRONI, A. DE TIENNE, V. COLAPIETRO, NÖTH W. *Peirce and Image 2. Working Papers and pre-publications*, Urbino, Università di Urbino "Carlo Bo" Edizioni.
- SEARLE J. (1969), *Speech Acts*, Cambridge, Cambridge University Press.
- TURING A. (1950), "I-Computing Machinery and Intelligence", *Mind*, 236: 433–460.
- VIOLI P. (2014), *Paesaggi della memoria. Il trauma, lo spazio, la storia*, Milano, Bompiani.
- WALSH L. (2014), "Tricks, Hockey Sticks, and the Myth of Natural Inscription: How the Visual Rhetoric of Climategate Conflated Climate with Character". In Birgit Schneider, Thomas Nocke (eds.) *Image Politics of Climate Change. Visualizations, Imaginations, Documentations*.

CONCLUSIONI

TRA SEMIOTICA E INTELLIGENZA ARTIFICIALE IL TERRENO DI UN CONFRONTO INTERDISCIPLINARE

ANTONIO SANTANGELO*

ENGLISH TITLE: *Between semiotics and artificial intelligence. For an interdisciplinary confrontation*

ABSTRACT: this article aims to draw some conclusions regarding the contents of the book *Semiotica e intelligenza artificiale* (Leone, Santangelo, eds., 2022). Furthermore, it attempts to identify some issues, regarding the relationship between semiotics and the various disciplines that deal with AI, that would allow research in this field to be relaunched. The main topics it deals with are twofold: the definition of the concept of intelligence starting from the faculty of correctly interpreting things, and the meaning that the design of the various artificial intelligence tools we use in our everyday lives takes on today.

KEYWORDS: artificial intelligence, semiotics, interdisciplinarity, interpretation, narratives, ideology, ethics, fairness

1. Sulla semiotica dell'intelligenza artificiale: alcune conclusioni e qualche rilancio

Questo breve testo prende le mosse da tutti quelli che lo hanno preceduto, all'interno di questo volume intitolato *Semiotica e intelligenza artificiale*. L'obiettivo è tentare di tirare le somme di quanto vi è stato scritto ed eventualmente rilanciare, in modo che il lettore possa individuare con più chiarezza cosa c'è ancora da studiare e come lo si debba fare, quando ci si accinge a collocarsi al crocevia tra la cosiddetta

* Università degli Studi di Torino.

“scienza dei segni” e le tante discipline che si occupano di IA.

Al di là della struttura tripartita del libro — che tiene conto dei problemi teorici che la semiotica pone a chi lavora nell’ambito dell’intelligenza artificiale e viceversa; si concentra su come la semiotica stessa possa essere utilizzata per implementare macchine “intelligenti”; e, infine, riflette su come si pensi e si parli, oggi, di questo genere di strumenti —, è chiaro che il dialogo con gli ingegneri, gli informatici, i filosofi, e tutti coloro che lavorano sull’IA si possa sviluppare su due piani: quello della definizione, teorica ma anche operativa, della nozione di intelligenza, che in semiotica ha a che vedere innanzitutto con la capacità di interpretare le cose, e quello di ciò che significa, a livello culturale, progettare e realizzare un certo tipo di tecnologia.

Sul primo punto, molto è stato scritto in queste pagine, problematizzando il concetto di “azione intelligente”. Senza ripetere le considerazioni dei vari autori i cui lavori compongono questo libro, si può comunque osservare che se si guarda a un testo classico, di natura tecnica, sullo sviluppo dell’IA, come quello di Russell e Norvig (2010), si vede chiaramente che essa viene concepita come uno strumento capace di agire razionalmente. Una macchina, infatti, è considerata intelligente, se persegue “il concetto ideale di intelligenza, che noi chiamiamo razionalità. Un sistema è razionale se, date le sue conoscenze, fa la cosa giusta” (Russell, Norvig 2010: 4). E ancora: “per ogni possibile sequenza di percezioni, un agente razionale dovrebbe scegliere un’azione che massimizzi il valore atteso della sua misura di prestazione, date le informazioni fornite dalla sequenza percettiva e da ogni ulteriore conoscenza dell’agente” (Russell, Norvig 2010: 47). Ma è evidente che questa forma di razionalità “economicistica”, volta a perseguire risultati pratici e utilitaristici, non è l’unica di cui è dotato un essere umano, come sottolineano giustamente alcuni filosofi della mente che si occupano di intelligenza artificiale (Lieto 2021) e come è risaputo in semiotica, dove un modello delle logiche di interpretazione del reale e di azione al suo interno, come quello di Floch (1990), mostra chiaramente che si può essere anche “utopici” o “ludici”, massimizzando “misure di prestazione” decisamente diverse.

Questa problematica è ben rappresentata in un film di fantascienza di M. Night Shyamalan, *After Earth* (2013), in cui Cypher e Kitai, un

padre e il suo giovane figlio, a causa di un'avarìa alla loro astronave, precipitano in un pianeta Terra divenuto inospitale, perdendo il pericoloso carico che trasportavano, vale a dire un mostro sanguinario che uccide inesorabilmente tutte le persone che dimostrino di temerlo. Cypher è famoso proprio per essere un grande giustiziere di queste bestie, perché non prova alcuna paura nei loro confronti, ma purtroppo è ferito gravemente e deve mandare Kitai in missione, per cercare di contattare qualcuno che possa giungere a salvarli. Si serve, quindi, dell'aiuto dell'intelligenza artificiale, per analizzare i dati del contesto in cui il ragazzo deve muoversi, i parametri vitali di quest'ultimo, le capacità degli strumenti tecnici di cui è dotato. In questo modo, gli indica cosa deve fare e quali scelte deve compiere per raggiungere il suo obiettivo. Ma, ad un certo punto, diventa evidente che, per colpa del mostro che gli sta dando la caccia, Kitai non ha alcuna speranza di riuscita e che l'unica cosa razionale da fare è cercare un rifugio, attendendo la morte: questo è il duro responso dell'IA e Cypher ordina a suo figlio di ritirarsi. Solo che quest'ultimo è animato dal senso di colpa per non aver saputo difendere, in passato, sua sorella, uccisa a sua volta da una delle bestie con cui sta combattendo, e vuole redimersi agli occhi di suo padre. Quindi agisce "irrazionalmente" e continua la sua disperata avventura. Ma proprio in quel momento, un uccello rapace che egli aveva aiutato, difendendo i suoi piccoli dall'attacco di un predatore, decide di sacrificare la propria vita per ricambiare il suo gesto altruistico e gli consente di arrivare alla trasmittente con cui può comunicare la sua richiesta di aiuto, sconfessando i calcoli dell'intelligenza artificiale. Quest'ultima, infatti, non poteva prevedere un gesto simile, perché non era programmata per pensare che si potesse agire in funzione di valori non utilitaristici.

Questo "apologo" può sembrare poco significativo, data la natura estrema delle vicende che racconta, ma proprio per la sua forma caricaturale mostra con estrema chiarezza una delle problematiche più evidenti poste dallo sviluppo attuale dell'intelligenza artificiale, vale a dire il fatto che, essendo uno strumento realizzato per consentire ai suoi finanziatori di raggiungere i propri obiettivi, ed essendo questi ultimi di natura spesso economicistica, essa viene addestrata a leggere la realtà secondo i dettami di una razionalità utilitaristica, volta a "massimizzare il valore atteso della sua misura di prestazione". Una misura che molto

raramente è incentrata sul perseguimento di obiettivi come il sacrificio per gli altri, perseguito dal giovane protagonista di *After Earth*.

Se questo accostamento appare ancora azzardato, si consideri ciò che si evince dallo studio, condotto da chi scrive insieme a Beretta, Lepri, Vetrò e De Martin (2019), sui parametri utilizzati per programmare algoritmi “*fair*”, giusti, che trattino le persone in modo da non discriminarle, quando operano delle scelte che le coinvolgono. Ciò che si evidenzia è che su sedici formule matematiche a cui si è soliti ricorrere a questo scopo, ben quattordici portano avanti criteri di giustizia “liberale” (Bobbio 1995) e solo due di giustizia “distributiva” (Rawls 1971). Questo significa che la forma di razionalità utilizzata per massimizzare la giustizia algoritmica è volta, nella maggior parte dei casi, a fare in modo che tutti vengano trattati dalle macchine alla stessa maniera, quando è necessario, per esempio, decidere se riconoscere loro un prestito bancario, una borsa di studio universitaria o la libertà vigilata, a prescindere dalla rispettiva appartenenza etnica, dal genere sessuale, eccetera. Ma ciò non tiene conto del fatto che non tutti vengono trattati in maniera eguale dalla società, prima di venire a contatto con le stesse macchine e che, dunque, in qualche caso, sarebbe necessario che chi è privilegiato rinunciassi ai propri vantaggi, redistribuendoli, appunto, a chi non ne ha. Per esempio, ci si potrebbe domandare se in un mondo del lavoro fortemente sbilanciato a favore degli uomini, a parità di curriculum, quello di una donna debba essere scelto per primo, da un algoritmo dotato di intelligenza artificiale con la funzione di scremare le migliaia di domande di assunzione giunte a una grande azienda, per offrire alla sua autrice un impiego e produrre una perequazione. Questo sembrerebbe contraddire una logica freddamente economicistica, in funzione di un'altra, in un certo senso, più “altruistica”. Ma questo modo di pensare è fortemente minoritario nell'ambito della programmazione dei sistemi informatici — dotati di IA e non solo — su cui si fonda la nostra vita di tutti i giorni, con tutte le conseguenze che si possono immaginare (Santangelo 2020 e 2022).

Questo genere di ragionamenti ci induce a riflettere anche sul secondo grande tema relativo al rapporto tra semiotica e intelligenza artificiale di cui si è scritto sopra, vale a dire quello del significato che deve assumere, oggi, un certo tipo di implementazione delle macchine

“intelligenti”. È evidente, infatti, che dietro alle convinzioni dei loro programmatori, ma anche dietro a quelle dei filosofi politici o morali a cui essi spesso si rivolgono, ci sono visioni del mondo diverse, fondate su valori e, quindi, su narrazioni differenti, in qualche caso di natura ideologica. Mettere in luce come funzionano e formalizzarle in una maniera tale che sia chiaro a tutti il senso che esse assumono, nel contesto culturale in cui viviamo, è un lavoro fondamentale, dato che questi strumenti sono e saranno sempre più diffusi, diventando una specie di “infrastruttura” della nostra società. Prima che i loro automatismi scompaiano ai nostri occhi e che ci abituiamo ad accettare le loro scelte, è necessario che ne comprendiamo la portata.

Ma anche tornando sul primo problema teorico di cui si scriveva sopra, a proposito del dialogo tra semiotica e intelligenza artificiale, vale a dire quello di come rendere davvero intelligente una macchina, non è possibile dimenticare le narrazioni a cui si è appena fatto cenno. Infatti, come si è anticipato, il tipo di intelligenza di cui si occupano semiologi e semiotici è quella interpretativa, che come si è visto nella definizione di Russell e Norvig è anche alla base del funzionamento delle macchine dotate di IA. Queste ultime, nelle intenzioni dei loro creatori, devono saper interpretare correttamente i dati che ricevono dall'esterno e, in funzione di queste interpretazioni, devono agire o suggerire azioni a chi le utilizza. Ma come dimostra Prieto (1975), in un saggio ormai divenuto classico in semiotica, il modo in cui gli esseri umani interpretano le cose dipende da ciò essi vogliono fare con queste ultime, dunque dalla loro idea pregressa di come funziona il contesto in cui operano e di ciò che possono ottenere, facendo determinate cose. Lo stesso avviene con l'intelligenza artificiale, il cui meccanismo “che massimizzi il valore atteso della sua misura di prestazione, date le informazioni fornite dalla sequenza percettiva e da ogni ulteriore conoscenza dell'agente” (Russell, Norvig *op. cit.*), dipende dal fatto che qualcuno che ha una certa visione del mondo e che ritiene di potervi o doverti agire in un certo modo, l'abbia programmata affinché “ritenga” che ciò che si trova davanti è significativo, al fine di raggiungere determinati risultati.

La letteratura sulla IA è piena di studi su come le scelte di interpretazione dei dati compiute a priori dai programmatori delle macchine, in funzione dei loro obiettivi pratici, si rivelino problematiche. Dai casi

di cui si è già scritto, in cui gli autori degli algoritmi di selezione del personale non tengono conto del fatto che le donne debbano essere viste come una categoria di persone da tutelare, a quelli sui sistemi informatici che sovrastimano la possibilità che un individuo di colore reitri un reato, o che sottostimano la facoltà di ripagare un debito da parte di chi proviene da certi contesti socioculturali (O'Neil 2016), si evince in maniera evidente che ciò che gli strumenti dotati di intelligenza artificiale ritengono significativo per operare la loro lettura della realtà dipende da una visione di quest'ultima molto precisa, volta a favorire un certo funzionamento dell'IA e, di conseguenza, del nostro mondo, che su di essa si appoggia.

In tutti questi casi, comunque, è dato per scontato che la macchina "intelligente" sappia riconoscere correttamente ciò che si trova a dover interpretare (il fatto di essere di fronte a una donna, a una persona povera o di colore, eccetera). Il problema è che qualcuno le insegna che le caratteristiche di ciò che ha davanti sono più o meno significative, in funzione degli obiettivi che deve raggiungere. Ci sono altri frangenti, però, in cui risulta evidente come l'IA non sappia quali sono gli elementi dotati di senso della realtà in cui opera, dimostrandosi, così, ben poco dotata di intelligenza.

Uno di questi ambiti è, per esempio, quello dei sistemi di raccomandazione di testi narrativi, come quelli che, sulle piattaforme di streaming audiovisivo, consigliano ai loro utenti quale film o serie tv guardare, sulla base di raffinati calcoli operati dall'intelligenza artificiale. È esperienza comune che questi strumenti non funzionino molto bene. Essi si basano molto su dati secondari, che indicano loro che persone classificabili nello stesso modo per via di una serie di parametri che le accomunano, hanno visto determinate cose, ma non sanno capire quale sia la ragione per cui le abbiano guardate, perché faticano a comprendere il significato di quei contenuti. Del resto, il lavoro interpretativo di un oggetto complesso come un prodotto audiovisivo non è lo stesso che viene richiesto per rendersi conto di trovarsi di fronte a un uomo, una donna, un cane, un gatto, una parola pronunciata da qualcuno, eccetera, come invece viene spesso domandato all'IA. Per questa ragione, in questi casi, l'intelligenza artificiale agisce di concerto con gli esseri umani, che interpretano per lei i film o le serie tv, descrivendoli

con etichette che la dovrebbero aiutare. Ma questi strumenti non sono sempre utili, poiché è possibile dimostrare, come chi scrive ha tentato di fare (Santangelo, Botta 2020; Ferraro, Santangelo, Botta 2021), che le persone che li producono non prestano attenzione a molti degli elementi che, in semiotica, vengono ritenuti come i più rilevanti, per dire qual è il significato di un testo narrativo.

Questo, comunque, non significa che la semiotica possa aiutare a risolvere tutti i problemi interpretativi dell'intelligenza artificiale, né che i suoi strumenti siano sempre migliori di quelli utilizzati in questo ambito. Così come viene affermato da molti degli autori di questo libro che qui si sta cercando di concludere, infatti, è risaputo che, molto spesso, le macchine sono in grado di produrre interpretazioni corrette delle cose e addirittura, a volte, sono capaci di proporre alcune a cui gli esseri umani non sanno arrivare, non possedendo le stesse facoltà di calcolo e di analisi dei cosiddetti "big data". In questi casi, può essere molto interessante confrontarsi con gli ingegneri e gli informatici, per capire come fanno a modellizzare, per mezzo della matematica e della statistica, il processo di assegnazione di significato al mondo. In questo modo, anche la semiotica può pensare di migliorare la propria comprensione di questi meccanismi, magari provando a compararli con quelli umani.

Come si evince dagli esempi sin qui riportati, però, questo tipo di lavoro richiede competenze interdisciplinari. Nell'articolo citato sopra, sui criteri di giustizia utilizzati nell'ambito della cosiddetta "*fairness* algoritmica", i co-autori della ricerca, insieme a chi scrive, erano un'economista specializzata in statistica con un dottorato in ingegneria informatica, due ingegneri informatici e un filosofo. Allo stesso modo, non è un caso che, in questo libro, Galofaro, che è un semiologo, abbia scritto con un matematico come Toffano, Bellentani con una linguista computazionale come Dall'Acqua, e che tutti gli altri autori facciano parte di gruppi di ricerca misti, in cui le competenze semiotiche si mescolano con quelle di chi si occupa di IA sotto molti altri aspetti, tra cui, preponderante, è quello tecnico. Senza questo modo di lavorare, è difficile produrre ricerche e risultati davvero sensati. Ciò che è auspicabile, dunque, è che questa attitudine all'interdisciplinarietà continui a essere perseguita.

Riferimenti bibliografici

- BERETTA, E., SANTANGELO, A., LEPRI, B., VETRÒ, A. E DE MARTIN, J.C., “The invisible power of fairness. How machine learning shapes democracy”, in *Proceedings of the 32nd Canadian Conference on Artificial Intelligence*, Springer, 2019, pp. 238–250.
- BOBBIO, N., (1995) *Eguaglianza e libertà*, Einaudi, Torino
- FLOCH, J.M. (1990) *Sémiotique, marketing et communication. Sous les signes, les stratégies*, Presses Universitaires de France, Paris
- LIETO, A. (2021) *Cognitive Design for Artificial Minds*, Routledge, London
- FERRARO G., SANTANGELO A. E A. BOTTA (2021) *Il significato di un suggerimento di visione. Riflessioni semiotiche sul sistema di raccomandazione di Netflix*, DigitCult Vol. 6, pp. 37–50
- O’NEIL, K. (2016) *Weapons of math destruction. How big data increases inequality and threatens democracy*, Penguin, New York
- PRIETO, L. (1975) *Pertinence et pratique*, Minuti, Paris
- RAWLS, J. (1971) *A theory of justice*, Harvard University Press, Cambridge
- RUSSELL, S. E P. NORVIG (2010) *Artificial Intelligence. A Modern Approach 3rd Edition*, Pearson Education Inc., publishing as Prentice–Hall; trad. it. (2010) *Intelligenza artificiale. Un approccio moderno. Terza edizione*, Pearson Prentice–Hall
- SANTANGELO, A. (2020) *Equità degli algoritmi e democrazia*, DigitCult Vol. 5, pp. 21–30
- . (2022) *Equità degli algoritmi, vita quotidiana e democrazia*, in M. DURANTE E U. PAGALLO, a cura di, *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*, Meltemi, Roma, pp.183–202
- SANTANGELO A. E A. BOTTA (2020) *Raccomandazioni sociosemiotiche. Considerazioni sul significato dei suggerimenti di visione del sistema di raccomandazione di Netflix*, DigitCult Vol. 5, pp. 53–64

AUTORI

Federico Bellentani è assegnista di ricerca post-doc all'Università di Torino, nel progetto ERC FACETS (Face Aesthetics in Contemporary E-Technological Societies), *Head of Marketing & Communications* in Injenia, azienda ICT con un forte background in semiotica, specializzata in progetti di innovazione che sfruttano le potenzialità dell'intelligenza artificiale e del *machine learning*, e Vicepresidente dell'International Association for the Semiotics of Space and Time. I suoi interessi di ricerca spaziano dalla semiotica della cultura alla tecnologia digitale, dalla geografia culturale al design degli spazi urbani. Ha conseguito il dottorato di ricerca alla School of Planning and Geography, Cardiff University (2017) e la laurea magistrale in semiotica (2013) presso l'Università di Bologna. Nel 2015-2016 è stato *visiting researcher* presso il Dipartimento di Semiotica dell'Università di Tartu. Ha pubblicato un libro, oltre quindici articoli e ha presentato la sua ricerca in numerose conferenze internazionali. Tra queste, ha tenuto un keynote in un convegno sul tema degli *hybrid security threat* in un panel insieme alla Presidente dell'Estonia Kersti Kaljulaid.

Anna Dall'Acqua è ricercatrice per FEM – Future Education Modena, centro di ricerca internazionale per l'innovazione in campo educativo. Si laurea all'Università di Bologna in Scienze Linguistiche (2020), con una tesi sulla progettazione e lo sviluppo di un agente conversazionale

in italiano. Dopo aver lavorato in una azienda emergente nell'ambito della *Conversational AI* come Conversation Designer, vince un assegno di ricerca al Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna, sullo sviluppo di metodologie linguistico-computazionali per la creazione e la progettazione di sistemi di dialogo. La sua ricerca accademica è affiancata dal lavoro di *Conversation Designer* in Injenia, azienda ICT specializzata in progetti di innovazione che sfruttano le potenzialità del *machine learning* in sinergia con approcci di linguistica e semiotica. I suoi interessi di ricerca spaziano dall'analisi della conversazione e la pragmatica del dialogo a strategie didattico-pedagogiche di glottodidattica e insegnamento di discipline linguistico-letterarie anche tramite approcci computazionali. Ha partecipato alla Summer School di LITHME–Language in the Human–Machine Era e ha tenuto dei moduli sulla progettazione di flussi conversazionali all'interno del Master in Digital Technology and Information Management (2019–2020, 2020–2021) e dell'Open Program in Artificial Intelligence and Machine Learning for Business (2020–2021, 2021–2022) presso la Bologna Business School. I risultati delle sue ricerche sono stati pubblicati sull'Italian Journal of Computational Linguistics e presentati in alcuni convegni nazionali.

Guido Ferraro, ha insegnato *Semiotica, Teoria della narrazione e Semiotica dei consumi e degli stili di vita* all'Università di Torino, dove è stato anche Direttore del *Master in progettazione e management del multimediale*. È stato inoltre Presidente dell'Associazione Italiana Studi Semiotici e del Centro Ricerche Semiotiche di Torino. Svolge attualmente ricerche sui modi in cui i diversi soggetti sociali formulano rappresentazioni della realtà contemporanea e delle possibili evoluzioni future. Per la semiotica e per la teoria della narrazione ha proposto un percorso di profondo rinnovamento, noto come prospettiva “neoclassica”, i cui risultati sono presentati in particolare nei libri *Fondamenti di teoria sociosemiotica* (2013), *Teorie della narrazione* (2015), *Semiotica 3.0* (2019).

Tulio Ferreira Leite da Silva is a Brazilian researcher at University of São Paulo (USP) where he is dedicated to natural language processing

and social media data mining. Silva has Discursive Semiotics as his epistemology and Computational Linguistics, Machine Learning, and Human–Computer Interaction as his working area. He aims to develop Glossematic Engineering, a metasemiotic applied to improving smart systems that use sentiment analysis and other text mining methodologies. He coordinates SemioCom (the experimental study group in Computational Semiotics at USP), whose current objective is to build the “Polimetro”, a sensor of political values on Twitter. His Ph. D. thesis’ theme is *Semiotic Intelligence and Coronavirus: Deep Learning and Discourse Analysis on Brazilian Portuguese Twitter*, which is sponsored by São Paulo Research Foundation (FAPESP).

Francesco Galofaro is associate professor at the Università IULM of Milan. He is former research fellow in the New Models of Sanctity in Italy ERC research project. He is an expert of Quantum Semantics, Religious Discourse, and Ethnosemiotics. Among his publications: *Eluana Englaro, la contesa sulla fine della vita*, Meltemi (2009); *Morphogenesis and Individuation*, Springer (2014, con A. Sarti e F. Montanari); *Dopo Gerico: i nuovi spazi della cura psichiatrica*, Esculapio (2015), *Il senso della tecnica: saggi su Bachelard*, Esculapio (2017, with P. Donatiello and G. Ienna). *Apprendisti mistici: Padre Pio e Ludwig Wittgenstein*, Mimesis (2022).

Gianmarco Thierry Giuliana (San Cataldo, 1989) è un semiologo esperto di videogiochi, VR e cultura digitale. Ha studiato Comunicazione presso l’Università di Catania, ha poi conseguito la laurea magistrale in Semiotica presso l’Università di Bologna e ha ottenuto il dottorato di ricerca in Semiotica e Media presso l’Università di Torino. Attualmente è assegnista di ricerca presso il Dipartimento di Filosofia e Scienze dell’Educazione di UniTo all’interno del progetto ERC FACETS (P.I. Massimo Leone) e docente a contratto per STUDIUM (UniTo). Specializzatosi sul tema del rapporto fra esperienza e interpretazione nelle realtà ludico–virtuali, ha scritto numerosi articoli scientifici su riviste di game studies, semiotica e filosofia. È stato relatore in conferenze e seminari tenutisi in Argentina, Belgio, Bulgaria, Cina, Finlandia, Francia, Germania, Grecia, Italia, Romania e Svezia.

Ha recentemente curato un libro sul tema della realtà estesa (*Meaning-Making in Extended Reality*, Aracne, 2020). Infine, ha collaborato sia come ricercatore che come game designer e divulgatore con enti privati ed associazioni come Club Silencio, Edugamers, EventHorizon, Forwardto, Horizon Psytech, VR Express.

Remo Gramigna è nato a Cosenza nel 1981. È assegnista di ricerca Post-doc FACETS (*Face Aesthetics in Contemporary E-Technological Societies*) presso il Dipartimento di Filosofia e Scienze dell'Educazione dell'Università degli Studi di Torino. Laureato in Scienze della Comunicazione presso l'Università "La Sapienza" Roma e in Semiotica a Tartu, in Estonia. Consegue il titolo di Dottore di Ricerca in Semiotica e Studi della Cultura presso l'Università di Tartu con una tesi sul problema filosofico del segno e della menzogna in S. Agostino. È stato *Visiting Scholar* presso l'Università di Siena, *Research Fellow* presso l'Università di Tartu e redattore della rivista internazionale di semiotica *Sign Systems Studies*. Ha pubblicato numerosi saggi su riviste e pubblicazioni nazionali e internazionali, soprattutto sulla storia della semiotica. Si è interessato ai problemi di semiotica generale, di semiotica della cultura, di teoria dei linguaggi e dei testi, di semiotica della manipolazione e dell'inganno. Ha pubblicato fra l'altro: *Augustine's theory of signs, signification, and lying* (De Gruyter, 2020); *The place of language among sign systems: Juri Lotman and Émile Benveniste* in "Sign Systems Studies", 2013; *Imagining others. Deception, Prediction, and Disguised Intentions in Strategic Interactions* in "Versus", 2020; *Le forme della maschera. Aspetti semiotici della manipolazione del volto e della plasticità dell'apparenza* in "Lexia, 2022.

Massimo Leone è Professore Ordinario di *Filosofia della Comunicazione, Semiotica della Cultura e Semiotica Visiva* presso il Dipartimento di Filosofia e Scienze dell'Educazione dell'Università di Torino, Professore part-time di *Semiotica* presso il Dipartimento di Lingua e Letteratura Cinese dell'Università di Shanghai (Cina), Membro Associato di Cambridge Digital Humanities, Università di Cambridge, e direttore del Centro di Studi Religiosi della Fondazione Bruno Kessler di Trento. È stato visiting professor in diverse università dei cinque continenti. È

autore di quindici libri, ha curato più di cinquanta volumi collettivi e pubblicato più di cinquecento articoli in semiotica, studi religiosi e studi visivi. È il vincitore di un ERC Consolidator Grant 2018, il più prestigioso finanziamento di ricerca in Europa. È caporedattore di *Lexia*, la rivista di semiotica del Centro di ricerca interdisciplinare sulla comunicazione dell'Università di Torino, caporedattore di *Semiotica* (De Gruyter), e direttore delle collane "I Saggi di Lexia" (Roma: Aracne), "Semiotics of Religion" (Berlino e Boston: Walter de Gruyter), e "Advances in Face Studies" (Londra e New York: Routledge).

Antonio Santangelo è professore associato presso l'Università degli Studi di Torino, dove insegna Semiotica, Semiotica delle culture digitali e Progettazione e management del multimedia per la comunicazione. È membro del gruppo di ricerca che lavora al progetto ERC denominato FACETS (Face Aesthetics in Contemporary E-Technological Societies), il cui *principal investigator* è il prof. Massimo Leone. È altresì *senior research fellow* del Centro Nexa su Internet e Società del Politecnico di Torino. I suoi principali interessi di studio sono legati allo sviluppo di una semiotica intesa come scienza sociale della significazione e alla teoria della narrazione, che applica in diversi campi: dai media studies, alla bioetica, alle indagini sul significato degli strumenti digitali nella nostra società. In particolare, sul tema dell'intelligenza artificiale, è stato uno dei curatori del libro bianco dell'Agid (Agenzia per l'Italia Digitale), intitolato *L'intelligenza artificiale al servizio del cittadino* (Aa.Vv. 2018), nonché l'autore di diversi articoli su libri e riviste, italiane e internazionali, dedicati al problema di come rendere l'IA capace di produrre interpretazioni raffinate di testi narrativi e a quello delle ideologie dominanti nell'ambito della *fairness* algoritmica.

Elsa Soro is a Faculty member at the Barcelona Tourism Hospitality and Gastronomy School of the University of Barcelona (CETT_UB) and former post-doc researcher at the University of Turin. She obtained her Ph.D. in Science of Language and Communication (Autonomous University of Barcelona) after which she combined professional activities as a consultant for private foundations and public bodies related to Tourism and Cultural Industries with lecturing in *Semiotics*, *Tourism*

and *Marketing* at different private and public universities. More info at <https://unito.academia.edu/ElsaSoro>

Simona Stano is Associate Professor at the University of Turin (UNITO, Italy). She was awarded a Marie Curie Global Fellowship for a research project (COMFECTION, 2019–2021) on the semiotic analysis of food communication. She also worked as Senior Researcher at the International Semiotics Institute (2015–2018) and as Visiting Researcher at New York University (NYU, US — 2019–2021), the University of Toronto (UofT, Canada — 2013), the University of Barcelona (UB, Spain — 2015–2016) and Observatorio de la Alimentación (ODELA, Spain — 2015–2016).

Dr. Stano deals mainly with semiotics of culture, food semiotics, body semiotics, and communication studies, and has published several papers, edited volumes (including special issues of top semiotic journals such as *Semiotica* and *Lexia*), and monographs (*I sensi del cibo*, 2018; *Eating the Other. Translations of the Culinary Code*, 2015) on these topics. She has presented many papers at national and international conferences, also organising and directing a number of scientific events and research projects. In recent years Dr. Stano has also collaborated with several universities in Italy and abroad. Moreover, she has delivered semiotic and cultural analysis for international agencies and private organisations.

Zeno Toffano is Professor at CentraleSupélec Université Paris Saclay and member of the CNRS–L2S Laboratory (Laboratoire des Signaux et Systèmes). He was trained as a Physicist at Université d'Orsay and has a PhD from the CEA (Commissariat à l'Énergie Atomique) in Saclay. His research experience has led him from NMR technology in superconductors, to Lasers and Optoelectronics and lastly Quantum Computing. He is an expert in Optical Communications and Quantum information. His actual interests go from quantum logic to quantum communications. He proposed a new method in logic: Eigenlogic based operators in Hilbert space. He developed an original method based on quantum correlations for Information Retrieval and investigated an original model for Cybernetic Agents: Braitenberg vehicle quantum

robots. He supervised 10 PhD-students and is the co-founder of the Master program “Optical Communications and Networks from Physics to Information Theory”, at U. Paris-Saclay. He coordinates research projects in CentraleSupélec on quantum information and quantum computing. He teaches: *Quantum Physics*, *Quantum Information*, *Statistical Physics*, *Optoelectronics* and *Telecommunications*.

Ugo Volli è professore di *Semiotica e Filosofia della Comunicazione*, in pensione con il titolo di professore onorario dell’Università di Torino. Nel 2010 ha ricevuto una laurea honoris causa dalla New Bulgarian University. Insegna e tiene conferenze in numerose università italiane e straniere. Ha al suo attivo circa 300 pubblicazioni scientifiche e oltre una quarantina di libri. Collabora con giornali, radio e televisioni. Ha svolto attività di consulenza per diverse aziende e istituzioni pubbliche e diretto numerose ricerche in ambito semiotico, fra cui un Prin sul palinsesto della città. Ha insegnato in numerose università italiane e straniere fra cui a lungo a Brown University e New York University. I suoi campi di ricerca principali: filosofia della comunicazione, teoria semiotica, analisi semiotica dei testi canonici della cultura religiosa ebraica, analisi della comunicazione politica, gusto. Fra i suoi libri più recenti: *Lezioni di filosofia della comunicazione* (Laterza 2008), *Parole in gioco* (Compositori 2009), *Domande alla Torah* (L’Epos 2012), *Periferie del senso* (Aracne 2016), *Il resto è interpretazione* (Belforte 2019), *Le donne di casa Boschi* (Skirà 2020), *Discutere in nome del cielo* (con Vittorio Robiati Bendaud-Guerini 2021), *Mai più* (Sonda 2022), *Musica sono per me le tue leggi* (Nave di Teseo 2022).

Cristina Voto è assegnista di ricerca all’interno del progetto ERC FACETS (Face Aesthetics in Contemporary E-Technological Societies) dell’Università di Torino, dove insegna *Comunicazione visiva* presso il corso di Laurea magistrale in Comunicazione e Culture dei media. Inoltre insegna *Strutture visive per le arti elettroniche* e co-dirige un progetto di ricerca presso l’Universidad Nacional de Tres de Febrero in Argentina. È membro del consiglio scientifico del Dottorato in Design e creazione dell’Universidad de Caldas, in Colombia, e curatrice della Biennale dell’immagine in movimento di Buenos Aires. Si occupa di

culture visive, arte digitale e studi sul futuro da una prospettiva intersezionale e decoloniale. Nel 2023 sarà *Visiting Scholar* presso la Vrije Universiteit, Amsterdam, per lo sviluppo di un progetto di ricerca sulla rappresentatività delle identità non-binarie nei sistemi di IA. È autrice della monografia *Monstruos audiovisuales. Agentividad, movimiento y morfología* (Aracne, 2021) e ha pubblicato articoli in riviste accademiche italiane e internazionali.

I SAGGI DI LEXIA

1. Gian Marco DE MARIA (a cura di)
Ieri, oggi, domani. Studi sulla previsione nelle scienze umane
ISBN 978-88-548-4184-0, formato 17 × 24 cm, 172 pagine, 11 euro
2. Alessandra LUCIANO
Anime allo specchio. Le mirouer des simples ames di Marguerite Porete
ISBN 978-88-548-4426-1, formato 17 × 24 cm, 168 pagine, 12 euro
3. Leonardo CAFFO
Soltanto per loro. Un manifesto per l'animalità attraverso la politica e la filosofia
ISBN 978-88-548-4510-7, formato 17 × 24 cm, 108 pagine, 10 euro
4. Jenny PONZO
Lingue angeliche e discorsi fondamentalisti. Alla ricerca di uno stile interpretativo
ISBN 978-88-548-4732-3, formato 17 × 24 cm, 356 pagine, 20 euro
5. Gian Marco DE MARIA, Antonio SANTANGELO (a cura di)
La TV o l'uomo immaginario
ISBN 978-88-548-5073-6, formato 17 × 24 cm, 228 pagine, 15 euro
6. Guido FERRARO
Fondamenti di teoria sociosemiotica. La visione "neoclassica"
ISBN 978-88-548-5432-1, formato 17 × 24 cm, 200 pagine, 12 euro
7. Piero POLIDORO
Umberto Eco e il dibattito sull'iconismo
ISBN 978-88-548-5267-9, formato 17 × 24 cm, 112 pagine, 9 euro
8. Antonio SANTANGELO
Le radici della televisione intermediale. Comprendere le trasformazioni del linguaggio della TV
ISBN 978-88-548-5481-9, formato 17 × 24 cm, 244 pagine, 19 euro
9. Gianluca CUOZZO
Resti del senso. Ripensare il mondo a partire dai rifiuti
ISBN 978-88-548-5231-0, formato 17 × 24 cm, 204 pagine, 14 euro
10. Guido FERRARO, Antonio SANTANGELO (a cura di)
Uno sguardo più attento. I dispositivi di senso dei testi cinematografici
ISBN 978-88-548-6330-9, formato 17 × 24 cm, 208 pagine, 13 euro

11. Massimo LEONE, Isabella PEZZINI (a cura di)
Semiotica delle soggettività
ISBN 978-88-548-6329-3, formato 17 × 24 cm, 464 pagine, 30 euro
12. Roberto MASTROIANNI (a cura di)
Writing the city. Scrivere la città Graffitiismo, immaginario urbano e Street Art
ISBN 978-88-548-6369-9, formato 17 × 24 cm, 284 pagine, 16 euro
13. Massimo LEONE
Annunciazioni. Percorsi di semiotica della religione
ISBN 978-88-548-6392-7, formato 17 × 24 cm, 2 tomi, 1000 pagine, 53 euro
14. Antonio SANTANGELO
Sociosemiotica dell'audiovisivo
ISBN 978-88-548-6460-3, formato 17 × 24 cm, 216 pagine, 14 euro
15. Mario DE PAOLI, Alessandro PESAVENTO
La signora del piano di sopra. Struttura semantica di un percorso narrativo onirico
ISBN 978-88-548-6784-0, formato 17 × 24 cm, 88 pagine, 9 euro
16. Jenny PONZO
La narrativa di argomento risorgimentale (1948–2011). Tomo I. Sistemi di valori e ruoli tematici. Tomo II. Analisi semiotica dei personaggi
ISBN 978-88-548-7751-1, formato 17 × 24 cm, 2 tomi, 788 pagine, 45 euro
17. Guido FERRARO, Alice GIANNITRAPANI, Gianfranco MARRONE, Stefano TRANI (a cura di)
Dire la Natura. Ambiente e significazione
ISBN 978-88-548-8662-9, formato 17 × 24 cm, 488 pagine, 28 euro
18. Massimo LEONE
Signatim. Profili di semiotica della cultura
ISBN 978-88-548-8730-5, formato 17 × 24 cm, 688 pagine, 40 euro
19. Massimo LEONE, Henri DE RIEDMATTEN, Victor I. STOICHITA
Il sistema del velo / Système du voile.
Trasparenze e opacità nell'arte moderna e contemporanea / Transparence et opacité dans l'art moderne et contemporain
ISBN 978-88-548-8838-8, formato 17 × 24 cm, 344 pagine, 26 euro
20. Mattia THIBAUT (a cura di)
Gamification urbana. Letture e riscritture ludiche degli spazi cittadini
ISBN 978-88-548-9288-0, formato 17 × 24 cm, 280 pagine, 20 euro

21. UGO VOLLI
Alla periferia del senso. Esplorazioni semiotiche
ISBN 978-88-548-9465-5, formato 17 × 24 cm, 380 pagine, 22 euro
22. GIAMPAOLO PRONI
La semiotica di Charles S. Peirce. Il sistema e l'evoluzione
ISBN 978-88-255-0064-6, formato 17 × 24 cm, 480 pagine, 22 euro
23. GUIDO FERRARO, ANTONIO SANTANGELO (a cura di)
I sensi del testo. Percorsi interpretativi tra la superficie e il profondo
ISBN 978-88-255-0060-8, formato 17 × 24 cm, 208 pagine, 12 euro
24. MARIANNA BOERO
Linguaggi del consumo. Segni, luoghi, pratiche, identità
ISBN 978-88-255-0130-8, formato 17 × 24 cm, 192 pagine, 16 euro
25. GUIDO FERRARO (a cura di)
Narrazione e realtà. Il senso degli eventi
ISBN 978-88-255-0560-3, formato 17 × 24 cm, 244 pagine, 15 euro
26. ALESSANDRO PRATO (a cura di)
Comunicazione e potere. Le strategie retoriche e mediatiche per il controllo del consenso
ISBN 978-88-255-0942-7, formato 17 × 24 cm, 164 pagine, 12 euro
27. VITALIANA ROCCA
La voce dell'immagine. Parola poetica e arti visive nei Neue Gedichte di Rilke
ISBN 978-88-255-0973-1, formato 17 × 24 cm, 176 pagine, 12 euro
28. VINCENZO IDONE CASSONE, BRUNO SURACE, MATTIA THIBAUT (a cura di)
I discorsi della fine. Catastrofi, disastri, apocalissi
ISBN 978-88-255-1346-2, formato 17 × 24 cm, 260 pagine, 18 euro
29. PATRÍCIA BRANCO, NADIRSYAH HOSEN, MASSIMO LEONE, RICHARD MOHR (edited by)
Tools of Meaning. Representation, Objects, and Agency in the Technologies of Law and Religion
ISBN 978-88-255-1867-2, formato 17 × 24 cm, 296 pagine, 18 euro
30. SIMONA STANO
I sensi del cibo. Elementi di semiotica dell'alimentazione
ISBN 978-88-255-2096-5, formato 17 × 24 cm, 228 pagine, 18 euro
31. GUIDO FERRARO
Semiotica 3.0. 50 idee chiave per un rilancio della scienza della significazione
ISBN 978-88-255-2318-8, formato 17 × 24 cm, 308 pagine, 18 euro

32. Simone GAROFALO
Narrarsi in salvo. Semiosi e antropo-poiesi in due buddhismi giapponesi
ISBN 978-88-255-2368-3, formato 17 × 24 cm, 516 pagine, 26 euro
33. Massimo LEONE
Il programma scientifico della semiotica. Scritti in onore di Ugo Volli
ISBN 978-88-255-2763-6, formato 17 × 24 cm, 228 pagine, 18 euro
34. Massimo LEONE, Bruno SURACE, Jun ZENG (edited by)
The Waterfall and the Fountain. Comparative Semiotic Essays on Contemporary Arts in China
ISBN 978-88-255-2787-2, formato 17 × 24 cm, 360 pagine, 25 euro
35. Jenny PONZO, Mattia THIBAUT, Vincenzo IDONE CASSONE (a cura di)
Languagescapes. Ancient and Artificial Languages in Today's Culture
ISBN 978-88-255-2958-6, formato 17 × 24 cm, 236 pagine, 22 euro
36. Andrea MAZZOLA
Trasumano mon amour. Note sul movimento H+ (scritti 2015-2019)
Prefazione di Riccardo de Biase
Traduzione di Annamaria Di Gioia, Federica Fiasca, Francesco Tagliavia, Giorgio Cristina
ISBN 978-88-255-3029-2, formato 17 × 24 cm, 288 pagine, 18 euro
37. Mattia THIBAUT
Ludosemiotica. Il gioco tra segni, testi, pratiche e discorsi
Prefazione di Ugo Volli
ISBN 978-88-255-3212-8, formato 17 × 24 cm, 236 pagine, 16 euro
38. Massimo LEONE
Colpire nel segno. La semiotica dell'irragionevole
ISBN 978-88-255-3381-1, formato 17 × 24 cm, 252 pagine, 18 euro
39. Massimo LEONE
Scevà. Parasemiotiche
ISBN 978-88-255-3455-9, formato 17 × 24 cm, 236 pagine, 16 euro
40. Federico BIGGIO, Victoria DOS SANTOS, Gianmarco Thierry GIULIANA (ed.)
Meaning-Making in Extended Reality. Senso e Virtualità
ISBN 978-88-255-3432-0, formato 17 × 24 cm, 336 pagine, 22 euro
41. Gabriele MARINO
Frammenti di un disco incantato. Teorie semiotiche, testualità e generi musicali
Prefazione di Andrea Valle
Postfazione di Ugo Volli
ISBN 978-88-255-3586-0, formato 17 × 24 cm, 244 pagine, 17 euro

42. Xianzhang ZHAO
Text – Image Theory: Comparative Semiotic Studies on Chinese Traditional Literature and Arts
ISBN 979-12-5994-008-7, formato 17 × 24 cm, 288 pagine, 22 euro
43. Cristina VOTO
Monstruos Audiovisuales. Agentividad, movimiento y morfología
ISBN 979-12-255-5994-419-1, formato 17 × 24 cm, 108 pagine, 10 euro
44. Silvia Barbotto, Cristina Voto, Massimo Leone (Editado por)
Rostrosferas de América Latina
ISBN 979-12-5994-921-9, formato 17 x 24 cm, 212 pagine, 18 euro
45. Jenny PONZO, Francesco GALOFARO (a cura di)
Autobiografie spirituali
ISBN 979-12-5994-878-6, formato 17 x 24 cm, 280 pagine, 16 euro
46. Massimo LEONE, Elsa SORO, Cristina VOTO (a cura di)
I cronotopi del volto
ISBN 979-12-218-0270-2, formato 17 x 24 cm, 260 pagine, 20 euro
47. Roberto FLORES ORTIZ
Magia Publicitaria. Semiótica de la eficacia simbólica
ISBN 979-12-218-0313-6, formato 17 x 24 cm, 184 pagine, 14 euro
48. Antonio SANTANGELO, Massimo LEONE (a cura di)
Semiotica e intelligenza artificiale
ISBN 979-12-218-0429-4, formato 17 x 24 cm, 308 pagine, 22 euro

Finito di stampare nel mese di gennaio del 2023
dalla tipografia «The Factory S.r.l.»
via Tiburtina, 912 – 00156 Roma