

6.

*LEGAL ANALYTICS* PER LA GIUSTIZIA TRIBUTARIA:  
LUCI E OMBRE DI UN'ESPERIENZA APPLICATIVA

di *Alessia Fidelangeli e Piera Santin* (\*)

SOMMARIO: 1. Introduzione. — 2. Il progetto Adele: *machine learning* e approcci logici, una prospettiva evolutiva. — 2.1. *Argument Mining*. — 2.2. *Outcome prediction*. — 2.3. Ontologia formale. — 2.4. *Citation Network Analysis*. — 3. Analisi delle criticità applicative: creazione del *dataset*. — 3.1. (segue) Disponibilità e qualità dei dati. — 3.2. (segue) Approcci metodologici — 4. I limiti giuridici: in particolare il problema dell'(in)spiegabilità. — 5. Spunti metodologici. — 6. Predizione e non perversione: l'uso dell'IA a supporto e non in sostituzione del giudice. — 7. Considerazioni conclusive.

1. *Introduzione*.

Con questo lavoro ci proponiamo di sviluppare alcune delle considerazioni che sono emerse a seguito del coinvolgimento nel progetto europeo e-Justice Adele.

L'obiettivo di tale progetto è applicare le tecnologie di intelligenza artificiale (IA), in gran parte riconducibili al filone della *legal analytics* (v. para. 2), a sentenze di merito in tre domini distinti, ovvero l'imposta sul valore aggiunto, i marchi e brevetti e gli aiuti di Stato di natura fiscale. Per ognuno dei primi due domini sono stati sviluppati due *dataset* a seconda che le sentenze fossero adottate da Corti italiane oppure bulgare. Per quanto riguarda gli aiuti di Stato, invece, si è fatto riferimento

---

(\*) Sono da attribuirsi ad Alessia Fidelangeli i paragrafi 1, 2, 2.2., 2.3., 3, 3.1.; a Piera Santin i paragrafi 2.1., 2.4., 3.2., 4, 6 e 7; il paragrafo 5 è frutto del lavoro delle due autrici.

alle sentenze d'appello della Corte di giustizia dell'Unione Europea (CGUE). Il progetto ha il duplice scopo di sviluppare una metodologia di estrazione della conoscenza in tali settori e di costruire uno strumento pilota con funzioni di supporto dell'attività giurisdizionale. La scelta dei diversi domini e delle diverse giurisdizioni è motivata dalla volontà di verificare se, ed eventualmente in che misura, le specificità dei diversi settori oppure delle tecniche di redazione e motivazione delle diverse Corti incidono sullo sviluppo di sistemi di IA.

Nel presente contributo ci soffermeremo innanzitutto sull'analisi delle funzionalità previste dal progetto e, in particolare, sull'*argument mining*, sull'*outcome prediction*, sullo sviluppo di un'ontologia formale e sulla *citation network analysis* (para. 2.). L'analisi è volta a ricostruire alcuni dei possibili utilizzi dei suddetti strumenti in ambito fiscale e a mostrare come l'intelligenza artificiale possa essere applicata in tale settore anche con funzioni diverse rispetto alla mera predizione. Alla luce di quest'analisi proporremo una disamina delle principali questioni applicative emerse nel progetto (para. 3), soffermandoci anche sull'aspetto, fondamentale, della spiegabilità (para. 4). Tali considerazioni saranno utilizzate per fornire, senza pretesa di completezza, degli spunti metodologici relativi all'ulteriore sviluppo dell'attività di ricerca scientifica nel campo dell'applicazione dell'IA al settore della giustizia tributaria (para. 5). Come si vedrà, lo stato dell'arte dell'IA nel settore giurisdizionale, sembra più adatto allo sviluppo tecnologie *a supporto* piuttosto che *in sostituzione* del giudice (para. 6).

Si ritiene, infatti, che nel settore dell'applicazione delle tecnologie di IA in ambito giuridico, in cui le competenze di giuristi e informatici debbono necessariamente coordinarsi, una riflessione che si fondi anche sull'esperienza applicativa possa essere utile per il dibattito scientifico sul tema. Si spera, inoltre, che alcune delle riflessioni proposte possano essere proficue per lo sviluppo di progetti futuri.

2. *Il progetto Adele: machine learning e approcci logici, una prospettiva evolutiva.*

Il progetto Adele si inserisce nel più ampio filone di ricerca sull'uso dell'intelligenza artificiale nel diritto. L'idea che questi due campi potessero essere uniti risale agli anni Settanta (1), sebbene la nascita di una comunità attiva di ricerca sull'IA e il diritto sia stata segnata dalla prima Conferenza internazionale sull'intelligenza artificiale e il diritto (ICAIL) (2), tenutasi nel 1987. Le prime ricerche si sono concentrate sull'esplorazione e sull'utilizzo di strutture logiche nell'argomentazione giuridica. Tuttavia, all'applicazione dell'intelligenza artificiale nel settore del diritto possono essere ricondotte moltissime attività basate sull'uso di tecnologie differenti.

Fino a pochi decenni fa i modelli di sistemi intelligenti più in uso erano riconducibili ai cd. *sistemi esperti*, cioè applicazioni informatiche che contengono rappresentazioni di conoscenze e competenze che si possono applicare, proprio come fanno gli esseri umani, per risolvere problemi, offrire consigli e svolgere una serie di altri compiti (3). Nei sistemi esperti la conoscenza giuridica viene assemblata all'inizio, secondo un approccio *top-down*, creando manualmente le regole del sistema per risolvere i problemi futuri (sistemi esperti basati sulla rappresentazione di regole) oppure identificando manualmente i fattori che hanno portato ad una certa determinazione giuridicamente rilevante in una situazione precedente e che potrebbero usati in casi futuri (sistemi esperti basati sulla

---

(1) 172B.G. BUCHANAN, T.E. HEADRICK, *Some Speculation About Artificial Intelligence and Legal Reasoning*, in *Stan. l. rev.*, 1970, vol. 40, n. 23 ipotizzarono tale relazione, proponendo una moltitudine di aree in cui l'IA avrebbe potuto contribuire al diritto.

(2) T. BENCH-CAPON et al., *A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law*, in *Art. intell. and law*, 2012, n. 20, 215-319.

(3) R. SUSSKIND, *The End of Lawyers? Rethinking the Nature of Legal Services*, Oxford University Press, 2010, 120 ss.

rappresentazione di casi) (4). I sistemi esperti applicano le regole o i fattori individuati ai casi nuovi e possono spiegare le scelte compiute in relazione alle regole o ai fattori applicati.

I sistemi esperti sono stati il principale paradigma di sistema intelligente fino agli anni Novanta e possono essere particolarmente utili nei settori in cui esistono regole connesse e precise, dal significato univoco (5). Per molto tempo l'approccio dominante alla modellazione computazionale del ragionamento giuridico è stato dunque di tipo *top-down* e ha fatto riferimento a questo paradigma. Tuttavia, i sistemi esperti possono essere utilizzati solo nei casi che ricadono nell'ambito applicativo delle regole precostituite, nel caso dei sistemi basati su regole, oppure che possono essere decisi sulla base dei fattori individuati in casi precedenti, nel caso dei sistemi basati su fattori. In entrambe le ipotesi, lo sviluppo del sistema richiede un lungo lavoro umano la cui correttezza potrebbe successivamente essere messa in discussione da mutamenti legislativi o giurisprudenziali (6). Infatti, la creazione, la modifica e l'aggiornamento delle regole in base all'evoluzione della legge richiedono sforzi umani e, dunque, costi elevati. Tale fenomeno è noto come “*bottleneck*” dell'acquisizione della conoscenza e ha limitato lo sviluppo dei sistemi esperti (7).

Le tecniche di elaborazione del linguaggio naturale, di apprendimento automatico e di estrazione di “significati” dalla previsione giuridica rispondono alla volontà di verificare se sia possibile individuare regole o generare modelli statistici da dati

---

(4) K.D. ASHLEY, *A Brief History of the Changing Roles of Case Prediction in AI and Law*, in *Law in Context*, 2019, n. 1, vol. 36, 95 parla di “formal representation of legal rules” e “representation of generalized case facts as factors”.

(5) G. SARTOR, A. SANTOSUOSSO, *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 2022, n. 7, 1759 ss.

(6) F. BEX, H. PRAKKEN, *The legal prediction industry: meaningless hype or useful development?*, 2020, <https://webspacescience.uu.nl/~prakk101/pubs/BexPrakkenAA2020English.pdf> (ultimo accesso: 26 ottobre 2023).

(7) K.D. ASHLEY, *A Brief History*, *cit.*, 94.

in entrata senza bisogno dell'intervento umano di modellazione esplicita delle norme, per superare il *bottleneck* dell'acquisizione della conoscenza. Come si vedrà meglio più avanti, spesso ciò avviene a scapito della possibilità di spiegare le previsioni (v. para. 4). Le tecnologie di apprendimento automatico sono spesso tecnologie di *machine learning*. Secondo l'accezione più comune del termine, per *machine learning* si intende un programma che apprende dall'esperienza E in relazione ad alcune classi di compiti T e alla misura di prestazione P, se le sue prestazioni nei compiti in T, come misurate da P, migliorano con l'esperienza E (8). Nel *machine learning*, dunque, al sistema vengono forniti metodi di apprendimento piuttosto che conoscenza formalizzata. Utilizzando tali metodi, i sistemi di intelligenza artificiale "apprendono" automaticamente a svolgere specifiche funzioni, mediante l'estrazione o la deduzione di informazioni rilevanti dai dati in *input*.

È, inoltre, opportuno chiarire che l'apprendimento della macchina può avvenire principalmente in tre modi: apprendimento supervisionato (*supervised*), apprendimento per rinforzo (*reinforcement*) e apprendimento non-supervisionato (*unsupervised*) (9). Nell'apprendimento supervisionato la macchina apprende mediante supervisione, cioè attraverso una fase di istruzione o addestramento in cui le viene fornito un ampio insieme di esempi, ciascuno della quale combina la descrizione di un caso alla risposta corretta. Su questa base la macchina stessa costruisce un modello generale, applicabile anche a casi nuovi, parzialmente diversi. L'apprendimento per rinforzo è simile all'apprendimento supervisionato, nel senso che entrambi comportano addestramento mediante esempi. Tuttavia, nell'addestramento per rinforzo, il sistema non ha bisogno di un istruttore, poiché apprende dai risultati delle

---

(8) T. MITCHELL, *Machine Learning*, McGraw Hill, 1997.

(9) Per questa tassonomia e le relative definizioni si veda G. SARTOR, *L'intelligenza artificiale e il diritto*, Torino, 2022, 41.

proprie azioni, cioè dalle ricompense o penalità (punti guadagnati o persi) che sono assegnate in base alle risposte del sistema. Nell'apprendimento non-supervisionato, infine, il sistema intelligente impara senza ricevere istruzioni, né da fonti esterne (apprendimento supervisionato), né dai risultati delle proprie attività (apprendimento per rinforzo). Le tecniche dell'apprendimento non-supervisionato sono usate ad esempio per il raggruppamento (cd. *clusterizzazione*), cioè per riunire insieme di oggetti che presentano somiglianze o connessioni rilevanti (i documenti che riguardano gli stessi oggetti o problemi, le persone con simili caratteristiche, le parole che hanno simili significati o funzioni) (10).

La conseguenza dell'utilizzo del *machine learning* è l'inversione del paradigma tradizionale di apprendimento: al sistema sono fornite grandi quantità di dati per permettergli di imparare autonomamente a estrarre le inferenze statistiche necessarie per svolgere il compito. In breve, l'apprendimento automatico sta invertendo il processo *top-down* di modellazione della conoscenza giuridica, consentendo agli approcci *bottom-up* di acquisire le conoscenze necessarie per prevedere i risultati, se non per offrire spiegazioni giuridicamente intelligibili di tali previsioni.

Esistono varie tecnologie che possono essere ricondotte al *machine learning*. Una di esse è legata all'elaborazione del linguaggio naturale (cd. *natural language processing*). Con tale termine si fa riferimento a un insieme di tecniche di elaborazione del linguaggio naturale.

È evidente che né i sistemi esperti, né i sistemi di *machine learning* sono realmente in grado di comprendere il significato e la semantica di un testo, giuridico e non, perciò in entrambe le tecnologie il sistema ragiona in base alla forma, ai simboli rappresentativi del linguaggio, e non in base al significato dei

---

(10) *Ibid.*

termini (11). Ciò che differenzia il sistema esperto dai sistemi di *machine learning* è il modo in cui viene fornita la conoscenza necessaria all'operatività: nei sistemi esperti è fondamentale l'interazione nella modellazione delle regole tra esperti del dominio ed esperti della rappresentazione formale della conoscenza; nel *machine learning* la base di conoscenza è costruita direttamente dal sistema sulla base dei dati ai quali ha accesso.

Nel progetto Adele si applicano eminentemente tecnologie di *supervised machine learning* per svolgere attività di *outcome prediction* e *argument mining*, mentre per la *citation network analysis* e per lo sviluppo di un'ontologia formale si utilizzano approcci misti, le cui caratteristiche saranno analizzate più dettagliatamente nel prosieguo (12). Gran parte delle attività svolte possono essere ricondotte al più ampio settore della *legal analytics*, ovvero l'attività di NLP, *machine learning* e altre tecniche computazionali per estrarre automaticamente i significati da banche dati di sentenze e altri documenti con valore giuridico (13).

È bene sottolineare che nessuna delle tecniche di *machine learning* utilizzate nel progetto Adele è riconducibile al novero dei sistemi generativi, vale a dire che in nessuna delle nostre applicazioni (*argument mining*, *outcome prediction*, ontologie e *citation network analysis*) l'algoritmo viene utilizzato per ge-

---

(11) A maggior ragione ove si consideri che tutta la rappresentazione dell'apprendimento da parte della macchina è basata sulla proiezione vettoriale dei lessemi.

(12) Esse sono basate su statistiche, diverse dal *machine learning*, per la *citation network analysis*; su tecniche di rappresentazione logica dei concetti nello sviluppo di un'ontologia formale.

(13) La definizione di *legal analytics* più autorevole è fornita da K. ASHLEY, *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge University Press, 2017, che rinvia a D.M. KATZ, M. BOMMARITO, *Legal Analytics - Introduction to the Course*, 2014. Tra le attività svolte difficilmente può essere ricondotta alla *legal analytics* in senso stretto lo sviluppo dell'ontologia formale poiché essa è stata costruita in base alla legislazione e alle conoscenze di un esperto del settore, mentre le altre attività, pur nella vaghezza del termine possono esservi ricondotte.

nerare nuovo testo. Dunque, l'esperienza di cui si darà conto nulla ha a che vedere con la manifestazione attualmente più di successo mediatico dell'IA, cioè ChatGPT4 e tutte le sue derivazioni. Ciò non toglie che anche le tecnologie da noi utilizzate (soprattutto in ambito di *argument mining* e *outcome prediction*) siano precipuamente basate sul NLP. Di conseguenza, tutti i risultati prodotti si fondano su un meccanismo di funzionamento che non prende in considerazione i rapporti logici tra i segmenti di testo analizzati.

Questa considerazione offre una imprescindibile chiave di lettura di tutta la ricostruzione degli esperimenti svolti perché influisce direttamente su molti dei limiti applicativi e giuridici alla reale utilizzabilità di queste tecniche, come si vedrà più dettagliatamente nei paragrafi 3 e 4.

## 2.1. Argument Mining.

In termini generali, rientrano in quest'ambito gli studi che, non solo nel settore legale, hanno l'obiettivo di addestrare i *software* di intelligenza artificiale al riconoscimento di premesse, conclusioni e schemi argomentativi presenti all'interno di testi strutturati (14). Si tratta, come si è detto, di un'area di ricerca che, sotto il profilo informatico, è riconducibile al più ampio ambito dei sistemi di *natural language processing*: l'obiettivo, infatti, è il riconoscimento di dette strutture argomentative solo a partire dall'analisi del linguaggio naturale.

In un sistema supervisionato, qual è quello di Adele, è necessario, preliminarmente, che un esperto del dominio mar-

---

(14) M. GRABMAIR et al., *Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools*, in *Proceedings of the 15th international conference on artificial intelligence and law*, 2015, 69; R. MOCHALES PALAU, M. F. MOENS, *Argumentation mining*, in *Artificial Intelligence and Law*, 2011, vol. 19, 1, 1 ss.; V. NICULAE, J. PARK, C. CARDIE, *Argument Mining with Structured SVMs and RNNs* in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Vancouver, Canada, 2017, 985.



chi dei testi di esempio (15), il cosiddetto *dataset* di addestramento, con l'obiettivo di individuare correttamente le porzioni di testo al cui riconoscimento si vuole addestrare la macchina. Ciò si è tradotto, nel caso di Adele, nell'individuazione di premesse, conclusioni e schemi argomentativi, il che significa, anche, individuare le connessioni logiche, più o meno fallaci, su cui si fonda la tesi sostenuta nel testo.

Prima di descrivere brevemente come questo tipo di esperimenti siano stati svolti nel contesto del progetto è tuttavia opportuno fare una breve premessa storico-metodologica. Al momento dell'avvio di Adele la letteratura in ambito di *argument mining*, e specificamente quella relativa all'ambito giuridico, propendeva, nella metodologia, per un approccio basato esclusivamente sul linguaggio. L'idea era che si potesse addestrare una macchina al riconoscimento, anche, dei nessi logici che costituiscono l'architettura di ogni testo argomentativo solo sulla base delle parole utilizzate in maniera ricorrente e riconoscibile nel momento in cui si introduce, ad esempio, un rapporto di consequenzialità o di antinomia (16).

Alcuni studi più recenti (17), tuttavia, hanno evidenziato come un approccio meramente linguistico rischi di essere insufficiente e apodittico, poiché l'identificazione e il riconoscimento di strutture logiche non può prescindere dal supporto ritraibile dai cosiddetti sistemi esperti, che fondano il loro funzionamento sulla logica formale.

---

(15) Vale a dire operi sul testo in XML, direttamente o mediante l'uso di strumenti di maggiore usabilità a ciò preposto, scrivendo stringhe di testo *machine-readable*, che consentano di individuare e qualificare la porzione di testo cui si fa riferimento. Nel caso di Adele, per esempio, ogni periodo appartenente alla motivazione della sentenza veniva classificato come premessa, di fatto o di diritto, o come conclusione e venivano indicati gli attributi relativi alla relazione logica ivi espressa.

(16) Si veda, sul punto M. LIPPI, P. TORRONI, *Argumentation mining: State of the art and emerging trends*, in *ACM Transactions on Internet Technology (TOIT)*, 2016, vol. 16, n. 2, 2016, 1.

(17) J. RABELO, R. GOEBEL, M.-Y. KIM et al., *Overview and Discussion of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021*, in *The Review of Socionetwork Strategies*, 2022, vol. 16, 111-133.

Questo limite, come si vedrà più diffusamente nel terzo paragrafo, è emerso chiaramente anche dall'esperienza applicativa del progetto Adele, che però è figlio del suo tempo e dello stato dell'arte scientifico dell'epoca. La descrizione tanto dei risultati quanto del metodo non può prescindere da questa considerazione, e dall'analisi dell'obiettivo ambizioso che ci si era posto, cioè far emergere la struttura logica del discorso giuridico sulla base della sola analisi del linguaggio.

Come è intuibile, la natura stessa delle pronunce giurisdizionali le rende un settore adatto allo svolgimento di esperimenti di *argument mining*, in ragione sia della loro struttura, sia dei loro contenuti (18). In particolare, per quel che concerne il progetto Adele, e con riferimento ai temi tributari, si è lavorato alla marcatura di due distinti *dataset* (19):

- sentenza d'appello della Corte di giustizia UE in materia di aiuti di Stato di natura fiscale;
- sentenze delle Commissioni Tributarie (antecedenti alla riforma) Regionali e, in minima parte, Provinciali in materia di IVA.

Per entrambi i *dataset* la marcatura ha riguardato l'individuazione delle premesse, a loro volta distinte in premesse di fatto o di diritto, e delle conclusioni; degli schemi argomentativi utilizzati (20); nonché dei rapporti reciproci di supporto e attacco tra le premesse e con la conclusione di ciascuna catena argomentativa.

---

(18) Si vedano, per esempio, gli esperimenti condotti sul *dataset* composto da sentenze della Corte EDU N. ALETRAS, D. TSARAPATSANIS, D. PREOTIUC-PIETRO, V. LAMPOS, *Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective*, in *PeerJ Comput. Sci.*, 2016, vol. 2, 93, R. MOCHALES, M.F. MOENS, *Study on the structure of argumentation in case law*, in *Proceedings of the 2008 conference on legal knowledge and information systems*, 2008.

(19) Senza considerare, in quest'analisi, i dati relativi al *dataset* bulgaro, di cui ancora non abbiamo i risultati sperimentali.

(20) Gli schemi utilizzati sono estratti e adattati dal modello di D. WALTON, C. REED, F. MACAGNO, *Argumentation Schemes*, Cambridge, 2008.

Ove possibile, la motivazione è stata divisa in distinte catene argomentative, poste a supporto della conclusione relativa ai singoli motivi di appello (qualificata come *<find>* nel *tagging* strutturale, v. para. 2.2.) relativa a ciascun motivo (v. *<claim>*, para. 2.2.) presentato dalle parti a sostegno della loro domanda (v. *<req>* para. 2.2.). La scelta di individuare le singole catene, pur funzionale alla realizzazione di una marcatura coerente, è stata influenzata sul piano pratico dal fatto che il *tagging* in quest'ambito è partito dalle sentenze della CGUE. Esse hanno uno schema formale di organizzazione del testo piuttosto rigoroso, in cui la Corte si preoccupa di distinguere e analizzare separatamente i singoli motivi posti a sostegno della richiesta presentata dalle parti. La presenza di un testo così strutturato ha reso perciò coerente e, quasi, necessitata l'organizzazione della marcatura per singole catene. Viceversa, nel lavoro sulle sentenze di merito delle Commissioni Tributarie, la mancanza di un'organizzazione del testo altrettanto rigorosa ha reso, in alcuni casi, la distinzione tra le catene più complessa. Per questa ragione l'attività di marcatura ha subito delle "forzature" (21)

Come si è detto, l'approccio algoritmico utilizzato in Adele è quello di pura analisi testuale: ciò significa che la classificazione delle premesse, ma soprattutto dei rapporti logico-argomentativi, avrebbe dovuto emergere, e prima ancora, essere individuato dai marcatori, sulla base del solo elemento "letterale" senza ausilio di astrazione o contesto (22). Viceversa, la marcatura sarebbe inutile allo scopo di addestrare un

---

(21) In particolare, nell'individuare separate conclusioni, spesso meno esplicite di quanto non fossero quelle della CGUE che verbalizzava l'accoglimento o il rigetto di ciascun motivo; per questa ragione alcuni supporti logici tra premesse sono stati spesso "non-rappresentabili", nella misura in cui legavano premesse appartenenti a catene differenti.

(22) In particolare, abbiamo considerato due classificatori neurali, le reti basate su trasformatori e le reti di attenzione residua, che sono ampiamente adottate per la previsione dei link argomentativi. Inoltre, abbiamo utilizzato un classificatore casuale uniforme e un classificatore di maggioranza come standard di riferimento.

algoritmo il cui funzionamento è basato esclusivamente sul linguaggio, senza alcuna componente di formalizzazione logica.

Sebbene siffatto approccio abbia mostrato i limiti di cui si dirà al paragrafo successivo, è interessante notare come, all'esito degli esperimenti, ma già in fase di marcatura, sia emersa una differenza quasi incolmabile tra i due *dataset*. Da un lato, infatti, le sentenze della CGUE, che sono stilisticamente più omogenee e che hanno blocchi argomentativi che si ripetono (v. para. 2.4.), hanno consentito di avere risultati sperimentali abbastanza soddisfacenti (23). Al contrario, le sentenze di diritto interno sono caratterizzate da una grande disomogeneità stilistica e, spesso, anche da un'esigenza di sintesi che porta a non esplicitare tutti i passaggi dell'argomentazione, pur comprensibili sulla base del contesto. L'effetto, dal punto di vista sperimentale, è quello di una minore efficacia di apprendimento e, quindi, di risultati meno soddisfacenti.

È in ogni caso opportuno chiarire che l'uso applicativo di un software di *argument mining* è, come dice il nome stesso, quello di ritrarre informazioni da un testo già scritto, e specificamente quelle relative alla sua struttura argomentativa. Non si tratta, cioè, di una prospettiva in cui l'intelligenza artificiale si possa sostituire all'uomo, sia esso giudice o avvocato. Si potrebbe trattare, piuttosto, di uno strumento a supporto delle attività delle parti, o, anche, di verifica e validazione della struttura argomentativa del giudice (24).

---

(23) Per una rappresentazione più dettagliata dei risultati si consenta di rimandare a G. K et al., *Detecting Arguments in CJEU Decisions on Fiscal State Aid*, Conference Paper, ArgMining 2022: 9 Workshop on Argument Mining, October 2022.

(24) Sull'utilizzo dell'*argument mining* come "strumento di verifica", sebbene in ambito non giuridico si veda A. LYDOS, T. LAGKAS, P. SARIGIANNIDIS, K. BONTCHEVA, *The Evolution of Argumentation Mining: From Models to Social Media and Emerging Tools*, in *Inf. Process. Manage.*, 2019, vol. 56, n. 6.

## 2.2. Outcome prediction.

Occorre specificare sin da subito che il termine “*predizione*” in ambito giuridico è utilizzato per la previsione dell’esito di una sentenza non ancora scritta; in ambito informatico, invece, esso è eminentemente legato alla metodologia e alla terminologia usate in relazione al *machine learning* (25). Nel settore informatico la definizione si applica, dunque, a svariate attività, che vanno dall’identificazione dell’esito di una sentenza già pubblicata (26), alla loro classificazione in base all’esito, utilizzando informazioni testuali o di altro tipo ma escludendo il verdetto della sentenza stessa (27), all’effettiva predizione dell’esito di un caso futuro (28). Il *dataset* di addestramento, dunque, può consistere in un gruppo di sentenze già scritte (nelle prime due ipotesi) oppure di atti delle parti (nella terza ipotesi).

In Adele la previsione giuridica si declina come uno studio retrospettivo a carattere sperimentale. Esso è volto cioè a verificare la possibilità che il sistema preveda l’esito di deci-

---

(25) M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction of court decisions*, in *Artificial Intelligence and Law*, 2023, n. 31, 198.

(26) Uno dei lavori più noti rispetto a questa funzione è quello di Aletras sulle sentenze della Corte EDU (si veda N. ALETRAS, et al. *Predicting judicial decisions of the European Court of Human Rights*, cit.).

(27) Si vedano, ad esempio, i lavori di I. CHALKIDIS, M. FEGRAIOTIS, P. MALAKASIOITIS, N. ALETRAS, I. ANDROUTSOPOULOS, *LEGAL BERT: “preparing the muppets for court”*, in *Proceedings of the 2020 conference on empirical methods in natural language processing*, 2898-2904 e M. MEDVEDEVA, M. VOLS, M. WIELING, *Juri says: prediction system for the European Court of Human Rights*, in *Artif. intell. law*, 2020, n. 28, 337-266.

(28) Si veda M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction*, cit., propone un’interessante ricognizione dei progetti che hanno applicato il NLP alla predizione in ambito giudiziario dal 2015 al 2021 e ha affermato che essi possono essere ricondotti a tre macro-categorie: (1) *outcome identification*, cioè l’identificazione dell’esito di una sentenza già pubblicata; (2) *outcome-based judgment categorisation*, cioè la classificazione dei documenti in base al loro esito; (3) *outcome forecasting*, cioè la previsione vera e propria dell’esito di una sentenza.

sioni già pubblicate delle Commissioni Tributarie Regionali (29) e, dunque, non riguarda la previsione di casi futuri sulla base delle richieste delle parti o di altri elementi rilevanti. Per svolgere tale tipo di predizione si possono distinguere due approcci principali. Nel primo approccio l'addestramento e la previsione si basano su fattori estratti manualmente o automaticamente dai casi (30); nel secondo approccio si basano sul contenuto testuale delle decisioni (NLP) (31).

In Adele si è seguito il secondo approccio. Occorre specificare che ad oggi non esistono altri esperimenti di applicazione di tecnologie basate unicamente su *machine learning* al testo puro a scopo predittivo in ambito fiscale (32). Inoltre, anche al di fuori del settore fiscale, la maggior parte degli esperimenti che sono ricondotti al genere della predizione, riguarda in realtà l'identificazione dell'esito o la classificazione delle sentenze, piuttosto che l'integrale automatizzazione del sillogismo giuridico, per decidere una controversia o una sua fase (33). Dunque, il progetto non ambisce a sviluppare un sistema da implementare concretamente in sostituzione del

---

(29) Come si è già sottolineato, gli esperimenti erano applicati alle sentenze pubblicate prima della riforma.

(30) Tali fattori possono anche non essere correlati al merito delle decisioni come in tutte le situazioni in cui si utilizzano a scopo predittivo elementi quali la nazionalità o l'estrazione sociale delle parti, piuttosto che la nazionalità o l'età del giudice.

(31) K.D. ASHLEY, *A Brief History*, cit., parla di “rule induction and decision trees”, “prediction via case-based argument models”, “prediction via machine learning” e “prediction with substantive information extracted from case texts”.

(32) Tecnologie di *machine learning*, seppur differenti, a scopo predittivo in ambito fiscale sono state utilizzate in un progetto tedesco (B. WALT, G. BONCZEK, E. SCEPANKOVA, J. LANDTHALER F. MATTHES, *Predicting the Outcome of Appeal Decisions in Germany's Tax Law*, in AA.VV., *Electronic Participation. ePart 2017. Lecture Notes in Computer Science*, vol. 10429, Springer, 89-99).

(33) Si esprime in questi termini, sottolineando la circostanza che in molte delle declinazioni utilizzate per il concetto di giustizia predittiva la “predizione” non è completa o non esiste affatto F. FARRI, *La giustizia predittiva in materia tributaria*, in *Riv. dir. trib.*, supplemento *online*, 12 ottobre 2022.

giudice (34). Per tale ragione si è affermato che si tratti di uno studio a carattere puramente scientifico e sperimentale.

In particolare, l'obiettivo del progetto è determinare le correlazioni tra le domande delle parti e l'accoglimento o il rigetto da parte del giudice in sentenze, giova ripeterlo, già redatte. A tal fine si è utilizzato un approccio di apprendimento supervisionato, in cui al sistema sono stati forniti i frammenti di testo rilevanti insieme ai corrispondenti esiti (35). L'apprendimento supervisionato, infatti, è il metodo di addestramento più utilizzato nella predizione in ambito giuridico (36).

In Adele, dunque, la predizione riguarda l'esito delle singole domande di parte e non l'accoglimento/rigetto dell'appello nella sua interezza; essa ha ad oggetto sentenze già redatte dal giudice, pertanto la sua finalità è la pura ricerca scientifica per l'elaborazione di una metodologia, non per suggerire un certo comportamento alle parti piuttosto che al giudice; la predizione si basa sull'analisi puramente testuale della sentenza e non sull'estrazione automatica di fattori o su altri elementi come l'identità o l'estrazione economico-sociale delle parti o del giudice ecc. (37); essa non è volta a fornire predizioni spiegabili (v. para. 4.).

---

(34) Più specificamente, nella piattaforma finale del progetto Adele (non ancora ultimata) sarà presente una sezione in cui un soggetto potrà inserire una porzione di testo e, dunque, anche il testo corrispondente alle domande di parte, per prevedere l'esito delle suddette domande. Tuttavia, dal momento che l'addestramento del sistema è avvenuto su sentenze già scritte sembra ragionevole che le attività svolte dal sistema abbiano esiti migliori in caso di sentenze già scritte.

(35) Per un approfondimento sulla differenza tra apprendimento supervisionato, rinforzato e non supervisionato si vedano A. SANTOSUOSSO, G. SARTOR, *La giustizia predittiva*, cit.

(36) M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction*, cit., 198.

(37) Si vedano, ad esempio, N. ALETRAS, D. TSARAPATSANIS, D. PREOTIUC-PIETRO, V. LAMPOS, *Predicting judicial decisions of the European Court of Human Rights*, cit.; M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction*, cit. e A.H. ZHANG, J. LIU, N. GAROUPA, *Judging in Europe: Do Legal Traditions Matter?*, in *J. of Comp. Law & Econ.*, 2017, vol. 13, n. 4.

In un primo momento, si è proceduto alla selezione di sentenze dalle banche dati della giustizia tributaria nel settore specifico delle operazioni imponibili/esenti IVA. Le sentenze sono state selezionate sulla base della classificazione esistente nella banca dati senza verificare che esse fossero effettivamente pertinenti rispetto ai temi selezionati. Tale scelta è stata motivata dalla considerazione che, nell'ipotesi dello sviluppo di un sistema di predizione, sarebbe ragionevole che esso prendesse i dati in *input* proprio da tale banca dati, poiché è quella in cui ci sono più sentenze disponibili, facendo affidamento quindi sulla classificazione esistente. Una volta selezionate le sentenze, esse sono state digitalizzate perché gran parte dei file erano disponibili in formato PDF proveniente da una scansione (cd. *PDF-immagine*) piuttosto che dalla conversione di un *file* di testo e, dunque, non erano fruibili al sistema nel formato di origine. Gli errori immediatamente percepibili all'esito del processo di digitalizzazione sono stati corretti manualmente.

Una volta digitalizzati i file, si è studiata la struttura delle sentenze, individuandone gli elementi essenziali, e si è proceduto alla redazione di linee guida per la marcatura delle stesse. L'obiettivo delle linee guida era di elaborare una metodologia comune per la marcatura degli elementi ritenuti rilevanti ai fini della predizione, come le domande e gli argomenti delle parti, la motivazione e le conclusioni dei giudici, e il dispositivo della sentenza.

Occorre preliminarmente chiarire che uno degli scopi del lavoro, poi rivelatosi per certi aspetti un limite, era quello di sviluppare linee guida per la marcatura che fossero applicabili non solo a uno specifico dominio (l'imposta sul valore aggiunto, ad esempio) ma a tutti i settori oggetto di indagine nel progetto con aggiustamenti minimi. Sulla base delle linee guida, si è proceduto alla marcatura manuale delle sentenze del *dataset*, cioè all'apposizione da parte di un soggetto umano di *tag* che permettessero al sistema di riconoscere le porzioni



della sentenza (38). L'attività è stata svolta da soggetti diversi, giuristi esperti e non esperti del settore, per verificare che le linee guida fossero sufficientemente specifiche da garantire risultati sufficientemente omogenei anche nell'ipotesi di annotazione da parte di soggetti con esperienze differenziate.

A seguito dell'annotazione, tecnologie di NLP, *machine learning* e reti neurali (39) sono state applicate ai testi delle sentenze. L'obiettivo era verificare se, sulla base delle domande e degli argomenti delle parti, il sistema fosse in grado di predire l'esito delle singole domande e con che margine di errore. Dunque, le tecnologie di IA sono state applicate per identificare automaticamente le correlazioni tra il contenuto testuale delle decisioni e i loro esiti sulla base della struttura lessicale, sintattica e semantica delle decisioni stesse.

Ciò ha comportato che la disomogeneità strutturale, sintattica e lessicale delle sentenze, che caratterizza le decisioni in ambito fiscale, abbia inciso sui risultati del progetto. Infatti, è stato subito evidente che esistessero sentenze di lunghezza differente, più o meno argomentate, e che, a seconda del collegio giudicante, venisse dato spazio differente alla ricostruzione del fatto, del *thema decidendum* e alla motivazione del giudice. Com'è noto, infatti, le sentenze in materia fiscale

---

(38) Esempi di *tag* sono i seguenti: [...] <judoff>LA COMMISSIONE TRIBUTARIA REGIONALE DI XXX SEZIONE 2</judoff>; <court C="Coll">riunita con l'intervento dei Signori: <judge ID="Judge1" R="Pres">giu1 Presidente</judge>; <judge ID="Judge2" R="Rapp">giu2 Relatore</judge>; <judge ID="Judge3" R="Judge">giu3 Giudice</judge></court> [...] <dec ID="Dec2" G="2" E="1" O="Req2" D="Find1">Accoglie l'appello</dec>; <costdec ID="Costdec10" G="2" P="A|B" O="Costreq1|Costreq2" D="Costfind1">e compensa le spese.</costdec>.

(39) Per una spiegazione sui meccanismi di funzionamento fondamentali delle reti neurali si veda G. SARTOR, *L'intelligenza artificiale e il diritto*, cit., 29 che spiega che le reti neurali mirano a riprodurre la struttura e la dinamica del cervello piuttosto che una sua particolare funzione (il ragionamento) e l'apprendimento avviene fondamentalmente mediante la modifica della forza delle connessioni tra i neuroni, così da dare alla rete la capacità di reagire agli stimoli. L'elaborazione della conoscenza da parte di una rete neurale non è simbolica ma piuttosto distribuita.

spesso si limitano a richiamare o esporre in modo conciso le domande delle parti oppure l'argomentazione giuridica che motiva una certa decisione, altre volte si esprimono diffusamente su di esse.

Gli esiti, dunque, hanno sofferto i limiti dovuti alle caratteristiche dei dati in *input*. In particolare, si fa riferimento alla formulazione delle domande delle parti e dei motivi di ricorso come riportate nella sentenza di secondo grado (in modo estremamente conciso e senza la disponibilità degli atti delle parti) e all'eterogeneità delle sentenze, in cui le questioni di diritto sostanziale non erano necessariamente la parte essenziale della decisione e che sono state estratte facendo riferimento alla classificazione, non sempre specifica, delle banche dati della giustizia tributaria.

Tali elementi hanno portato all'ottenimento di risultati in linea con quelli ottenuti in progetti simili ma assolutamente insufficienti per immaginare un utilizzo in ipotesi applicative reali, anche volendo trascurare gli aspetti giuridici che a nostro parere precluderebbero la legittimità e l'opportunità di tale utilizzo (40). Si ricorda, inoltre, che gli esperimenti svolti hanno riguardato sentenze già scritte e, dunque, per valutare qualsiasi ulteriore considerazione applicativa occorrerebbe preliminarmente valutare l'adeguatezza e l'opportunità dell'uso della metodologia adottata per gli atti delle parti. Una linea di indagine futura potrebbe essere quella relativa alla combinazione di tecnologie diverse, come nel caso in cui si utilizzassero sia tecnologie di *machine learning* basate sul testo, sia l'estrazione di fattori.

---

(40) La previsione era corretta in una percentuale di casi che oscilla tra il 50 e il 78% a seconda dell'algoritmo utilizzato (si veda F. GALLI et al, *Predicting outcomes of Italian VAT decisions*, in: *Legal Knowledge and Information Systems: JURIX 2022*, IOS Press, 2022, 362, 188-193 (atti di: Jurix, Saarbrücken, 14-16 Dicembre 2022). Per gli esiti ottenuti da altri progetti in ambito giuridico, si rinvia a M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction*, cit.

### 2.3. *Ontologia formale.*

Per ontologia formale si intende una descrizione formale ed esplicita di un insieme astratto di oggetti, concetti, relazioni ed altre entità che appartengono a uno specifico dominio. L'ontologia viene definita *formale* in quanto la rappresentazione fornita è basata su simboli e quindi computabile (41). Il suo obiettivo è quello di gestire e rendere fruibili al sistema non solo la sintassi, ma anche il “significato” delle informazioni giuridiche, con il risultato ultimo di migliorare le potenzialità dei sistemi che si occupano di ragionamento automatico (42). Si è sostenuto che un'ontologia presenti cinque caratteri fondamentali: è *descrittiva* in quanto è una forma di rappresentazione della conoscenza; è *esplicita* in quanto contiene definizioni; è *astratta* in quanto fornisce modelli attraverso i quali esprimere la conoscenza, ma non la conoscenza stessa; si limita a un dominio, cioè a uno specifico settore della realtà visto da una certa prospettiva (43); è *formale* per le ragioni già analizzate.

Le ontologie sono indispensabili innanzitutto per la ricerca documentale. Ad esempio, in un sistema per la ricerca delle sentenze che disponga di un'ontologia formale che rappresenti i concetti principali e le loro relazioni, se si cercassero le sentenze relative alle operazioni imponibili IVA, si potrebbero ottenere anche i documenti che trattano delle sottoclassi del concetto di operazione imponibile. Nel nostro caso, ciò significherebbe, ad esempio, ottenere anche i documenti relativi ai concetti di cessioni di bene e delle prestazioni di servizi, in assenza di un riferimento esplicito all'“operazione imponibile” nel testo della sentenza. Inoltre, se il sistema fosse modellato in modo tale da sottolineare la relazione biunivoca tra il diritto di

---

(41) G. CONTISSA, *Le ontologie giuridiche e la loro integrazione con i sistemi basati su regole*, in *Quaderni della rivista di diritto privato*, Bari, 2017, 87-109.

(42) *Ivi*, 88.

(43) *Ivi*, 90.

un soggetto nei confronti di un altro e l'obbligo di quest'ultimo nei confronti del primo, se si chiedesse al sistema di trovare tutti i documenti in cui è statuito un obbligo di A nei confronti di B, si troverebbero anche le sentenze in cui si parla di un diritto di B nei confronti di A. Se già nei sistemi di ricerca tradizionali era frequente il ricorso a tassonomie, *thesauri* e elenchi di sinonimi, l'aspirazione ulteriore delle ontologie è di andare al di là del dato terminologico per fornire al sistema informatico una conoscibilità quanto più possibile vicina alla semantica del testo di riferimento (44).

Nel progetto Adele abbiamo costruito un'ontologia dei concetti principali relativi a un aspetto specifico della disciplina IVA (operazioni imponibili/esenti) con il duplice scopo di potenziare le attività di ricerca e di organizzazione della conoscenza giuridica e di verificare la possibilità di collegare in modo automatico le sentenze del dataset ai concetti più rilevanti del dominio. Dal momento che uno degli obiettivi dell'ontologia era il suddetto collegamento, l'ontologia ha riguardato unicamente i profili sostanziali affrontati nelle sentenze.

L'ontologia è stata sviluppata a partire da una mappa concettuale in cui sono stati inseriti i concetti rilevanti e le relazioni tra essi (45). Gli elementi costitutivi di un'ontologia formale sono, infatti, la *classi*, cioè i concetti generali del dominio di interesse (ad esempio, "esenzione"); le *relazioni*, cioè i rapporti tra una classe e l'altra (ad esempio, la relazione *subclass* tra le classi "operazione imponibile" e "cessione di

---

(44) *Ivi*, 91.

(45) Lo sviluppo di un'ontologia prevede varie fasi: in una prima fase ci si limita ad enumerare i concetti rilevanti del dominio sulla base dell'utilizzo che se ne vuole fare e dei destinatari. In una seconda fase, con un approccio *top-down* oppure *bottom-up*, si definiscono le classi fondamentali dell'ontologia e la gerarchia tra essi. Si strutturano poi le classi secondo una gerarchia tassonomica e si individuano gli attributi e le proprietà dei concetti. Infine, si popola l'ontologia con le istanze.

beni”); le *proprietà*, cioè le relazioni binarie che specificano le caratteristiche della classe (46).

Per i concetti inseriti nella mappa si è proceduto a individuare le definizioni, gli esempi, i sinonimi, i concetti correlati, nonché la legislazione europea e nazionale rilevante in cui i concetti sono menzionati o definiti. L’obiettivo era disporre anche degli elementi ulteriori che permettono di collegare una sentenza al concetto dell’ontologia, non solo, dunque, sulla base della presenza del lessema che individua la nozione nella sentenza stessa. Ad esempio, la presenza di definizioni, esempi e termini correlati permette di ricondurre al concetto di “cessioni di beni” le sentenze in cui si parla del trasferimento della proprietà di un oggetto specifico in assenza di qualsiasi riferimento esplicito al termine “cessione di beni”. Anche il riferimento alla disposizione rilevante oppure ai sinonimi del concetto possono essere elementi utili a suggerire un collegamento tra decisione e nozione.

Sulla base di questi elementi si è implementata l’ontologia con il linguaggio OWL e si sta attualmente sviluppando un algoritmo che, sulla base di tecnologie di similarità semantica, modelli statistici e reti neurali, proporrà i collegamenti tra concetti e sentenze. La correttezza degli *output* forniti dall’algoritmo sarà verificata sulla base dello svolgimento manuale della medesima attività e della comparazione (strumentale alla successiva correzione dell’algoritmo) dei risultati ottenuti.

La creazione dell’ontologia è stata particolarmente complessa per due ordini di ragioni. Il primo è un aspetto comune a tutte le ontologie che intervengono nel settore del diritto ed è legato alla natura del linguaggio giuridico. Esso è composto in gran parte di termini appartenenti alla lingua naturale e solo per una parte relativamente minore di termini tecnico-giuri-

---

(46) G. CONTISSA, *op. prec. cit.*, 93.

dici (47). Dunque, l'ontologia include anche nozioni appartenenti ad altri settori, concetti fattuali (ad esempio "sangue umano") (48) oppure mutuati dalle scienze sociali (ad esempio, "intenzione"). L'individuazione della definizione di tali nozioni non è sempre agevole.

Il secondo ordine di ragioni riguarda la scelta del settore dell'imposta sul valore aggiunto. Da un lato, la scelta di un settore in cui fosse presente una disciplina armonizzata era necessaria per permettere di avere una struttura comune alla quale ricondurre le sentenze del *dataset* sia italiano, sia bulgaro (v. para. 2). Dall'altro lato, però, com'è noto, l'attuazione della disciplina IVA, soprattutto nell'ordinamento nazionale, ha avuto luogo con riferimento a concetti e definizioni spesso diversi da quelli utilizzati nella Direttiva. Inoltre, spesso i concetti utilizzati nel dominio di riferimento non sono definiti né dalla legislazione europea, né da quella nazionale, come avviene in relazione ad alcuni dei concetti usati in materia di esenzioni.

Per quanto riguarda il primo ordine di questioni, per i concetti fattuali, si è fatto riferimento al significato di senso comune, ci si è basati sulle definizioni rinvenibili ad esempio nell'Enciclopedia Treccani. Per quanto riguarda il secondo ordine di problemi, in caso di definizioni diverse a livello europeo e nazionale, sono state inserite entrambe e si è data priorità alla definizione nazionale. In caso di divergenze tra i termini utilizzati per individuare una nozione (oltre che delle definizioni), i diversi lessemi utilizzati per individuare il concetto in ambito nazionale sono stati elencati tra i sinonimi. Nel caso, invece, in cui né la legislazione europea né quella nazio-

---

(47) M. JORI, A. PINTORE, *Introduzione alla filosofia del diritto*, Torino, 2014, 182.

(48) Sui concetti fattuali si veda T. SPAAK, *Explicating the Concept of Legal Competence*, in AA.VV., *Concepts in Law*, a cura di J.C. HAGE, D. PFORDTEN, Springer, 68 e 82.

nale definissero il termine, si è inserita la definizione fornita nella giurisprudenza della Corte di giustizia.

Infine, per il collegamento dei concetti dell'ontologia alle sentenze del *dataset*, cioè l'attività ancora in corso di sviluppo, si può immaginare che una delle principali difficoltà sarà legata a quanto già evidenziato in relazione alla predizione. L'affidamento alla classificazione del sito della giustizia tributaria, non sempre corretta, e la circostanza che nelle sentenze siano spesso discusse in maniera preponderante questioni relative al procedimento, all'ammissibilità delle domande, al rispetto dei termini o alla congrua motivazione della sentenza di primo grado complicano l'individuazione del collegamento. A questo proposito, dunque, sarebbe interessante circoscrivere gli esperimenti a sentenze selezionate prestando particolare attenzione alla prevalenza dei profili sostanziali e alla correttezza della classificazione delle banche dati per valutare se e in che misura migliori il risultato.

#### 2.4. Citation Network Analysis.

Si tratta di una specifica applicazione al settore giurisdizionale di quegli studi statistici che prendono il nome di *network analysis*, la cui pietra miliare ha superato il limite del dibattito scientifico per approdare alla cultura pop nella forma della teoria dei sei gradi di separazione (49). A partire da un importante studio sulla giurisprudenza della Corte Suprema degli Stati Uniti d'America in materia di diritto all'aborto (50), nell'ultimo decennio è emerso un crescente interesse degli

---

(49) S. MILGRAM, *The small world problem*, in *Psychology Today*, 1967, n. 1, 61.

(50) J. H. FOWLER et al., *Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court*, in *Political Analysis* 15, 2007, 324.

studiosi e della letteratura scientifica in ordine all'uso della *network analysis* nell'ambito degli studi giuridici (51).

Questo tipo di indagine, che tende ad introdurre, anche in ambito giuridico, elementi di analisi quantitativa, ha conosciuto uno sviluppo, seppure limitato, con riferimento alla CGUE, la cui giurisprudenza è stata elaborata sia nella relazione con le conclusioni degli avvocati generali (52), sia nello sviluppo diacronico in temi specifici (53), sia, da ultimo, nell'ambito di un più ampio studio sull'uso del precedente (54). Mancano invece precedenti esperienze applicative riferibili a giudizi di merito di diritto italiano, ivi compresi quelli della giustizia tributaria.

In realtà, la *network analysis* ha sfruttato la tecnologia e la potenza di calcolo in rapido sviluppo per fornire metodi nuovi e atipici (le scienze giuridiche si basano su un approccio eminentemente qualitativo), nonché studi più trasparenti, sistematici, completi, riproducibili e falsificabili (55). Gli scienziati sociali si sono rivolti alla *legal network analysis*, ponendosi le domande più spinose sul diritto e sul comportamento giudiziario. Hanno potuto basare le loro indagini su insiemi di

---

(51) B. WELLMAN, S. D. BERKOWITZ, *Social Structures: A Network Approach*, Bingley, 1988, 5; per una ricostruzione più recente della letteratura esistente in materia, si veda S. P. BORGATTI et al., *Network Analysis in the Social Sciences*, in *Science*, 2009, n. 323, 892.

(52) J. FRANKENREITER, *Are Advocates General Political? An Empirical Analysis of the Voting Behavior of the Advocates General at the European Court of Justice*, in *Review of Law & Economics*, 2018, vol. 14, no. 1.

(53) See M. DERLÉN, J. LINDHOLM, *Characteristics of Precedent: The Case Law of the European Court of Justice in Three Dimensions*, in *German Law Journal*, 2015, 1073 ss.; ID., *Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments*, in *European Law Journal*, 2014, vol. 20, 667; M. DERLÉN et al., *Coherence out of Chaos: Mapping European Union Law by Running Randomly Through the Maze of CJEU Case Law*, in *Europarättslig Tidskrift*, 2013, 517.

(54) U. ŠADL and F. TARISSAN, *The Relevance of the Network Approach to European Case Law: Reflection and Evidence*, in C. KILPATRICK and J. SCOTT (eds.), *New Legal Approaches to Studying the Court of Justice*, first edition, Oxford 2020.

(55) *Ibid.*



dati completi e prontamente disponibili, su fonti di dati ufficiali e gratuite e su algoritmi e connessioni esplicite, piuttosto che su relazioni speculative.

All'interno del progetto di Adele, tuttavia, gli elementi di *network analysis* erano specificamente (e più semplicemente) tesi a verificare l'apparato di citazioni, sia giurisprudenziali sia legislative (56), rilevabile dalle sentenze del *dataset*. Ciò al duplice scopo di consentire l'individuazione dei precedenti di maggior peso e di poter accedere a una rappresentazione grafica delle reti. Le tecnologie utilizzate, e in particolare il ricorso a espressioni regolari, sono quindi riconducibili al più ampio settore dell'NLP, su cui intervengono però, in questo caso, anche elementi di logica, tesi a rendere più preciso il risultato dell'estrazione.

Sotto questo aspetto è emersa una significativa differenza tra i due *dataset* di riferimento. Come noto, lo stile redazionale delle sentenze della CGUE prevede che le citazioni dei precedenti, salvo rari casi, siano formulate con specifico riferimento all'esatto paragrafo del precedente citato (57). Questo contenuto sostanziale è pedissequamente rappresentato dallo schema di marcatura con cui le sentenze in formato XML

---

(56) L'analisi di rete, sebbene più diffusamente indagata con riferimento ai precedenti giurisprudenziali, può essere effettuata con riferimento a qualsiasi sistema complesso di contenuti, dunque anche nell'analisi delle citazioni legislative.

(57) Si veda ad esempio la citazione proveniente dalla pronuncia Commissione / Polonia, C-562/19: "First of all, it should be noted that, according to the settled case-law of the Court of Justice, action by Member States in areas that are not subject to harmonisation by EU law are not excluded from the scope of the provisions of the FEU Treaty on monitoring State aid (see, to that effect, judgment of <REF.DOC.ECR><DATE ISO="20060622">22 June 2006</DATE>, <NAME .CASE><HT TYPE="ITALIC">Belgium and Forum 187</HT> v <HT TYPE="ITALIC">Commission</HT></NAME.CASE>, <NO.CASE>C-182/03 and C-217/03</NO.CASE>, <NO.ECLI ECLI="ECLI:EU:C:2006:416">EU:C:2006:416</NO.ECLI>, paragraph <REF.NP.ECR REF="NP0081">81</REF.NP.ECR></REF.DOC.ECR>). The Member States must thus refrain from adopting any tax measure liable to constitute State aid that is incompatible with the internal market".

vengono messe a disposizione del pubblico (v. para. 3.2.). Ciò significa che, per ogni citazione, è stato possibile risalire esattamente non solo alla sentenza, bensì addirittura al paragrafo cui si faceva riferimento. Non solo, in ragione della scalabilità (58) di questa struttura e della reperibilità delle sentenze, ci è stato possibile procedere a una ricostruzione regressiva delle citazioni, fino ad individuare la prima occorrenza di un determinato precedente, e a poter verificare quali, nel rimando delle citazioni incrociate, siano le pronunce più significative, con riferimento al complesso apparato di citazioni in *subiecta materia*, non solo al dataset originario.

Questa ricostruzione più articolata, cioè arricchita dalla possibilità di far regredire la ricerca oltre il primo livello di citazioni, non è stata possibile con riferimento alle sentenze di merito in materia di imposta sul valore aggiunto. In questo caso, infatti, il limite tecnico che abbiamo dovuto superare era duplice: da un lato, lo stile redazionale delle sentenze italiane di merito non prevede un riferimento al precedente tanto preciso quanto quello utilizzato dalla giurisprudenza europea (né una formula omogenea di citazione delle sentenze), né, *a fortiori*, offre una marcatura effettuata alla fonte; d'altro canto, la reperibilità delle sentenze (perlopiù di Cassazione), citate dai giudici di merito era decisamente limitata, oltre che resa complessa dalla disomogeneità degli stili di citazione. Per ovviare a questa difficoltà, quindi, è stata svolta un'attività manuale di individuazione dei *pattern* di citazione delle sentenze e sono state adeguate le espressioni regolari per l'individuazione delle diverse tipologie.

---

(58) Il concetto si riferisce alla capacità di un sistema di ospitare un numero crescente di elementi o di oggetti, di elaborare volumi di lavoro crescenti con grazia, e/o di essere suscettibile di ampliamento. Quando si progetta un sistema, spesso si richiede che sia scalabile (A.B. BONDI, *Characteristics of scalability and their impact on performance*, in *Proceedings of the 2nd international workshop on Software and performance (WOSP '00)*. Association for Computing Machinery, New York, NY, USA, 195-203).

Inoltre, per escludere citazioni non riconducibili all'argomentazione del giudice senza necessità di aggiungere nuovi *tag*, si è deciso di limitare l'estrazione delle citazioni nel testo (che viceversa avrebbero dovuto essere marcate manualmente, con un dispendio di tempo ingiustificato) sulla base di criteri di selezione basati sulle marcature, rispettivamente, strutturali (*outcome prediction*) e argomentative (*argument mining*). In particolare, la ricerca doveva essere limitata alle sole citazioni presenti nella motivazione, vale a dire nella porzione di testo marcata come tale e, tra queste, sono state ulteriormente selezionate quelle rinvenibili all'interno di periodi classificati come premesse di diritto. In tal modo, si è evitato il rumore derivante dall'errata individuazione come citazione, ad esempio, dei riferimenti alla sentenza impugnata, i quali sono formalmente (lessicalmente) del tutto identici a una citazione inserita nel testo con una funzione logico-argomentativa.

L'opportunità e l'utilità di questo genere di analisi è risultato essere, quindi, grandemente influenzato dalla qualità e dalla omogeneità, in termini sia tecnici sia contenutistici, del *dataset* di partenza. L'indubbia qualità dei dati della CGUE in materia di aiuti di Stato, oltre alla loro maggiore omogeneità (59), ha consentito di ottenere risultati interessanti dal punto di vista sperimentale e analitico, che potranno essere base di partenza per ulteriori sviluppi. Solo a titolo esemplificativo, si può ricordare che l'analisi di rete ha fatto emergere con chiarezza alcuni specifici profili procedurali, e in particolare la distinzione fra interesse individuale e interesse diretto ad agire delle terze parti. Concetti che, in ragione della loro prossimità sul piano applicativo ed ermeneutico, erano sempre rimasti confusi e confondibili in un'area grigia dell'interpretazione europea. Si potrebbe altresì sottolineare come

---

(59) Si trattava di una selezione ridotta (i soli casi di appello in materia di aiuti di Stato di natura fiscale) e in quanto tale verificabile, mentre così non poteva certo dirsi per una materia varia e multiforme qual era quella dell'originaria selezione sull'IVA.

l'analisi abbia fatto emergere una centralità, o perlomeno una maggior occorrenza, di citazioni di precedenti connessi a questioni di natura procedurale, e non, come forse ci si sarebbe potuto aspettare, sostanziale. È, però, facile l'obiezione per cui questo dato è una conseguenza diretta dell'aver scelto, per la creazione del *dataset*, sentenze di appello, il cui contenuto è, almeno in parte, influenzato dalla vicenda processuale, diversamente da quanto sarebbe potuto accadere con un *dataset* composto unicamente di rinvii pregiudiziali.

Questo è sicuramente vero, ma è una considerazione che attiene all'analisi qualitativa del settore, oltre che al tema della selezione del *dataset* come elemento qualificante della ricerca. Il risultato più importante di questa parte del progetto è piuttosto quello metodologico, non quello direttamente contenutistico. A parità di pulizia dei dati, quindi con tutta la giurisprudenza europea, ma, potenzialmente, con qualunque forma di sentenza resa disponibile in un formato XML marcato, questo metodo di analisi è ripetibile all'infinito, senza limiti tematici o di giurisdizione. Il che significa, o potrebbe significare, lo sviluppo di un sistema di ricerca e organizzazione della conoscenza e dei percorsi giurisprudenziali pregressi molto più articolato, raffinato e affidabile di quelli oggi a disposizione.

### 3. *Analisi delle criticità applicative: creazione del dataset.*

L'esperienza maturata nel progetto Adele ha portato alla luce una serie di considerazioni applicative e giuridiche che meritano di essere analizzate autonomamente. Alle considerazioni giuridiche relative all'applicazione dell'intelligenza artificiale in ambito fiscale si è dato ampio spazio nel dibattito scientifico recente (60). Diversamente, le considerazioni applicative sono state oggetto di scarsa attenzione.

---

(60) Si vedano, in termini generali, C. BUCCICO, *Giustizia predittiva e processo tributario: problemi e prospettive*, relazione al Convegno "La digitalizzazione del-

Occorre innanzitutto concentrarsi sulle questioni legate alla creazione del *dataset*. Nel corso del lavoro si sono applicate le tecnologie di IA a un settore ampio, facendo affidamento sulla classificazione già utilizzata nelle banche dati esistenti. L'adozione della classificazione proposta nelle banche dati nazionali, motivata dalle ragioni già analizzate (v. para. 2.2.), ha portato a includere nel *dataset* anche sentenze che non avevano effettivamente ad oggetto i profili sostanziali su cui ci si intendeva soffermare. Di conseguenza, malgrado le sentenze annotate fossero quantitativamente significative, quelle che avevano effettivamente ad oggetto la disciplina sostanziale delle operazioni imponibili/esenti era in realtà un numero inferiore. È stato sottolineato in più lavori che l'applicazione delle tecnologie a fini predittivi, a causa delle difficoltà di rappresentazione della conoscenza giuridica, dovrebbe generalmente coprire domini relativamente ristretti (61). Dunque,

---

l'Amministrazione finanziaria tra contrasto all'evasione e tutela dei diritti del contribuente", Napoli, 5 ottobre 2022; S. DORIGO, *Intelligenza artificiale e norme antiabuso: il ruolo dei sistemi "intelligenti" tra funzione amministrativa e attività giurisdizionale*, in *Rass. trib.*, 2019, 728 ss.; F. FARRI, *La giustizia predittiva*, cit.; B. KUZNIACKI, *The marriage of artificial intelligence and tax law: Past, Present, and Future*, in *Kluwer International Tax Blog*, January 25, 2019; S. HOFFER, *What if Tax Law's Future is Now?*, in *The Ohio State Technology Law Journal*, n. 1, vol. 16, 68-72; A. MARCHESELLI, *Intelligenza artificiale e giustizia predittiva: il bivio tra Giustiniano e il Leviatano e il pericolo Coca Cola*, in *Riv. dir. trib.*, supplemento online, 20 ottobre 2022; V. MASTROIACOVO, *Prevedibilità, predittività e umanità del giudicare in materia tributaria*, in *Riv. dir. trib.*, supplemento online, 14 febbraio 2023, 1; C. SACCHETTO, *Processo tributario telematico e giustizia predittiva in ambito fiscale*, in *Rass. trib.*, 2020, 41-54; A. VOZZA, *Intelligenza artificiale, giustizia predittiva e processo tributario*, in *Fisco (I)*, 2019, n. 32/33, 2019, 3154 ss. Con riferimento all'implementazione delle tecnologie di IA nell'ordinamento tributario italiano e allo specifico progetto Prodigit, le cui riflessioni possono però essere in parte estese anche a questioni di ordine generale, si veda il recente fascicolo monografico a cura di A. MARCHESELLI e E. MARELLO, *Loading Prodigit del diritto naturale al diritto digitale: l'intelligenza artificiale nella giustizia tributaria*, in *Riv. dir. trib.*, 2022.

(61) K.D. ASHLEY, *A Brief History*, cit., 94 sottolinea che, a causa delle difficoltà di rappresentazione della conoscenza giuridica, i modelli coprono generalmente ambiti relativamente ristretti, come il diritto dei segreti commerciali o il diritto dei proprietari e degli inquilini. Poiché ogni modello è una semplificazione,

l'inclusione di un'ampia varietà di sentenze potrebbe aver inciso in modo negativo sugli esiti della previsione. Ciò dimostra, in primo luogo, la centralità della preliminare e corretta classificazione dei dati ai fini dell'applicazione delle tecnologie di IA e suggerisce che, sebbene si tratti di un tema meno discusso della predizione, potrebbe essere utile investire tempo e risorse nel supporto delle tecnologie alla classificazione tassonomica. In secondo luogo, si potrebbe valutare, in prospettiva futura, di limitare l'annotazione a un numero inferiore di sentenze ma più omogenee dal punto di vista sia dei temi trattati, sia stilistico/strutturale (62).

Inoltre, l'inclusione di sentenze in cui si tratta diffusamente di aspetti procedurali o processuali non ha aiutato a collegare con precisione i concetti dell'ontologia alle sentenze del *dataset*. Per ovviare a tale difficoltà si potrebbe valutare alternativamente una selezione più precisa delle sentenze del *dataset* oppure l'ampliamento dell'ontologia o la creazione di ulteriori ontologie che abbiano ad oggetto anche aspetti ulteriori rispetto a quelli sostanziali.

Per quanto riguarda la difformità stilistica delle sentenze, anch'essa ha un impatto sulla capacità di apprendimento del sistema. Le tecnologie utilizzate, infatti, sono destinate a incontrare maggiori difficoltà di apprendimento tutte le volte in cui i dati di addestramento sono poco coerenti. Testi privi di omogeneità stilistica e contenutistica comportano l'impossibilità di individuare un *pattern* ripetuto e consolidato e, per l'effetto, l'incapacità di ripetere l'analisi su testi nuovi. Ad esempio, nel caso dell'*argument mining*, il sistema deve apprendere sulla base della sintassi delle frasi e della ricorrenza di determinate strutture linguistiche o espressioni, in particolare per quanto riguarda l'individuazione di un legame logico di

---

è importante definire un caso d'uso, cioè un'applicazione a cui il modello sarà destinato, per la quale il modello sia sufficientemente dettagliato e accurato.

(62) L. K. BRANTING, *Scalable and explainable legal prediction*, cit., 232.

attacco o supporto tra premesse. In questo scenario l'estrema difformità linguistica che caratterizza le sentenze redatte in ambito nazionale ha ostacolato la correttezza del lavoro. Anche per quanto riguarda la *network analysis*, la difformità delle modalità di citazione dei precedenti e della legislazione non favorisce la possibilità di estrazione automatica, effettuata sulla base della ricerca di formule lessicali, in assenza di un testo marcato. Se, infatti, la citazione di un precedente di Cassazione può apparire formulato in modi diversi (ad esempio, per indicare l'organo giudicante: C. Cass., Cass., Cassazione; per indicare il numero della sentenza: nn, aaaa, nn del aaaa, nn/aaaa, sentenza nn del gg mese aaaa, etc.), è impossibile essere certi di poter ritrovare automaticamente in un testo tutte le citazioni presenti. Se, viceversa, la citazione viene fatta in maniera rigorosa e sempre formalmente identica, il rinvenimento dei precedenti, pur in assenza di marcatura, è molto più semplice (com'è emerso dal lavoro sulla giurisprudenza europea (63)).

Inoltre, la diversa lunghezza delle sentenze in materia IVA ha reso spesso difficile identificare argomentazioni sufficientemente precise da poter essere annotate in modo univoco e rigoroso da un esperto del dominio. Ciò è stato evidente sia in relazione alla predizione, poiché spesso le sentenze non individuavano con sufficiente chiarezza le domande e gli argomenti delle parti, limitandosi a rinviare alla sentenza di primo grado oppure agli atti delle parti, ma anche nel caso dell'*argument mining*. In quest'ambito, infatti, ciò che è apparso evidente già in sede di marcatura è la tendenziale scarsa organicità delle motivazioni. Sebbene per un esperto della materia sia possibile comprendere, o almeno intuire, qual è il senso e quali

---

(63) Per una spiegazione più esaustiva dal punto di vista tecnico si rimanda a G. SARTOR et al., *Automated extraction and representation of citation network: a CJEU case-study*, in R. GUIZZARDI, B. NEUMAYR, (a cura di), *Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science*, 2023, vol. 13650. Springer, Cham.

sono le ragioni che hanno condotto il giudice a determinarsi in una certa direzione, è altresì vero che l'analisi di precisione, tesa alla frammentazione del testo e all'individuazione dei nessi logici lascia non poche perplessità: in primo luogo per il gran numero di "omissioni", cioè di premesse implicite, e in seconda battuta per una certa tendenza alla giustapposizione delle premesse, più che alla loro concatenazione.

### 3.1. (segue) *Disponibilità e qualità dei dati.*

Alle considerazioni appena evidenziate si aggiungono quelle specificamente legate alla disponibilità e qualità dei dati. Infatti, è noto che il risultato fornito dagli algoritmi è necessariamente influenzato dai dati in *input*, dei quali bisogna curare la qualità, l'indipendenza della fonte e l'accessibilità (64).

Lo svolgimento di attività nel settore della giustizia tributaria soffre innanzitutto i limiti legati all'accessibilità, poiché le sentenze di merito sono attualmente accessibili nelle banche dati pubbliche in forma anonimizzata solo in maniera limitata. Ad esempio, banche dati come "Leggi d'Italia" rendono disponibili solo una parte delle sentenze. Le sentenze sono invece disponibili in misura quantitativamente maggiore nella banca dati della giustizia tributaria. Tale disponibilità, però, presenta due limiti. Il primo è la classificazione imprecisa che si è già evidenziata. Il secondo riguarda la disponibilità in formato PDF-immagine che, come si è in parte anticipato, richiede il preliminare riconoscimento automatico del testo e tale attività difficilmente porta ad esiti precisi. Di conseguenza si deve accettare alternativamente di avere dati che presentino errori legati alla corretta formulazione dei termini oppure

---

(64) A. NATALE, *Introduzione. Una giustizia (im)prevedibile?*, in [https://www.questionegiustizia.it/rivista/articolo/introduzione-una-giustizia-imprevedibile-\\_572.php](https://www.questionegiustizia.it/rivista/articolo/introduzione-una-giustizia-imprevedibile-_572.php) (ultimo accesso: 26 ottobre 2023).



un'attività di correzione umana. La prima soluzione è difficilmente accettabile qualora la qualità del documento sia bassa. La seconda soluzione porta sicuramente a risultati più precisi e qualitativamente migliori, ma può richiedere una sproporzione tra il tempo impiegato per l'addestramento e i risultati ottenuti, in ragione della grande quantità di lavoro necessaria per correggere il testo delle sentenze.

Nel nostro caso, inoltre, l'attività di controllo umano, ha riguardato anche l'anonimizzazione delle sentenze. Infatti, a fronte di un quadro giuridico non chiarissimo per quanto riguarda la necessità dell'anonimizzazione e di tecnologie di anonimizzazione non gratuitamente disponibili al pubblico, si sono anonimizzate manualmente le sentenze. La qualità e la disponibilità dei dati, dunque, hanno un rilievo dirimente in qualsiasi attività che si proponga di applicare le tecnologie di intelligenza artificiale al diritto. Nella nostra esperienza, ad esempio, il lavoro di marcatura specifica, cioè connessa al progetto, e, successivamente, di analisi dei dati è stato molto più semplice nel caso delle sentenze europee.

Ciò è vero sia sotto il profilo dell'accessibilità, giacché tutte le sentenze pubblicate sono disponibili sui portali di Curia ed EUR-Lex, sia sotto quello della qualità dei dati. Le sentenze pubblicate negli ultimi dieci anni, infatti, sono rese disponibili dall'Unione Europea in formato XML originale, con la conseguenza di non richiedere alcun intervento di normalizzazione e correzione. Ne deriva che l'affidabilità del dato, tanto in *input* quanto in *output* è incomparabilmente maggiore, con una maggiore attendibilità anche dei risultati prodotti in sede sperimentale.

Non solo, le sentenze rese disponibili in XML sul sito di EUR-Lex sono anche già dotate in un proprio sistema di marcatura (lo schema Formex (65)) che, nel nostro caso

---

(65) Per una descrizione di questo schema si veda <https://op.europa.eu/it/web/eu-vocabularies/formex> (ultimo accesso: 26 ottobre 2023).

specifico, è stato utilizzato come base di conoscenza per l'effettuazione di tutti gli esperimenti in tema di *citation network analysis*. Siffatto formato, validato all'origine dal fatto che ne è autore e responsabile la stessa istituzione cui le pronunce sono riconducibili, ha reso possibile uno sviluppo più articolato, meno dispersivo e molto più omogeneo del settore di ricerca specifico.

Ciò concorre a dimostrare la grande importanza della qualità dei dati, allo scopo di raggiungere risultati quanto più possibile affidabili e, comunque, non alterati da interventi esterni. Inoltre, la possibilità stessa di lavorare su dati pronti ha un effetto positivo sul piano dell'efficacia e dell'organizzazione temporale del lavoro sia degli esperti giuridici, sia dei programmatori. Basti pensare che, sempre in tema di *citation network analysis*, il tempo impiegato a rendere le sentenze di merito italiane "lavorabili" dal punto di vista formale è stato molto significativo. Purtroppo, però, a un tale impiego di tempo non è corrisposto un risultato altrettanto soddisfacente, con l'aggravante della perdita di ogni elemento di ricorsività a causa del già ricordato problema della parziale inaccessibilità dei dati in materia tributaria (66).

### 3.2. (segue) *Approcci metodologici.*

Le considerazioni tecniche sui dati, sulla loro forma e accessibilità, non possono che cedere il passo a una seconda tipologia di riflessioni: quella metodologiche. L'approccio seguito nel corso del progetto, anche in conseguenza della sua natura pionieristica, è quello che si può riassumere nel concetto di *learning by doing*. Non c'è bisogno di soffermarsi

---

(66) Sebbene le sentenze citate fossero quasi esclusivamente sentenze di Cassazione, duole rilevare che anche l'accesso a queste ultime è problematico e limitato, sia dal necessario ricorso a banche dati proprietarie, sia dalla mancanza di un unico repository ufficiale cui poter avere libero accesso, almeno per ragioni di ricerca.

troppo ampiamente su quanto questo metodo sia antitetico rispetto al mondo degli studi giuridici e su quanto, quindi, abbia richiesto a tutti i soggetti coinvolti un lavoro per il reciproco adattamento.

La marcatura di un testo sulla base di etichette individuate a monte, in ragione della conoscenza specifica della materia, dunque astrattamente nel dialogo tra informatici e giuristi, è un'attività di natura ambigua, al contempo prescrittiva (anche se non all'apparenza) e descrittiva. La scelta di chi marca il testo di attribuire una determinata etichetta ad una porzione di testo comporta (prescrive), nell'ambito del sistema di riferimento per l'addestramento dell'IA, un'assunzione di significato di quella porzione di testo. Si tratta di una funzione "creativa" di significato, per cui, ad esempio, ai fini dell'apprendimento della macchina, una determinata stringa di testo (intrinsecamente inconoscibile al *software*) non è una premessa di fatto prima che il marcatore la qualifichi come tale, utilizzando a questo scopo un linguaggio comprensibile alla macchina stessa (quello dei *tag*). Peraltro, un'ulteriore dimostrazione della natura prescrittiva dei *tag* e, al contempo, della sua fallibilità, dovuta alla mediazione dell'umano, è il fatto che, in caso di marcature in doppio cieco del medesimo testo vi saranno molto probabilmente delle discrasie. Discrasie che dovranno essere risolte proprio per evitare che si crei confusione nell'applicazione delle regole di apprendimento (o, che è lo stesso, che vengano applicate regole diverse in condizioni analoghe).

Il testo marcato, vale a dire il testo in cui alle frasi (alle successioni di lettere) è stato attribuito un significato (il *tag*), diviene così un testo descrittivo per l'addestramento dell'intelligenza artificiale. Quando si addivene alla fase di *training*, in cui la macchina dall'osservazione del mondo come descritto dai *tag* dovrebbe apprendere delle regole per l'analogo riconoscimento su testi nuovi, la marcatura ha esaurito la sua funzione prescrittiva, creatrice del mondo, per approdare a quella descrittiva dell'esistente.

È pacifico, quindi, che sia particolarmente importante ai fini giuridici l'elaborazione di un *set* di marcatura sufficientemente preciso dal punto di vista giuridico, ma al contempo non tanto rigoroso da rendere impossibile adattarlo alla varietà di formule espressive presenti nei testi da marcare. Questo risultato non può che essere raggiunto per il tramite di progressivi adattamenti a partire dall'ipotesi di marcatura iniziale, per poter contemperare le due esigenze sulla base dell'esperienza maturata.

La procedura di revisione, adattamento e parziale riforma in corso d'opera (anche) delle categorie prescrittive (i *tag*) è certo uno dei profili metodologici di più difficile gestione per il giurista, orfano dell'affidabilità (per quanto sempre relativa) riconosciuta alla fattispecie. Di questa difficoltà di adattamento si è fatta diretta esperienza anche nel progetto Adele. Prima di raggiungere l'equilibrio finale, l'elaborazione del sistema di marcatura, ha, infatti, conosciuto fasi di eccessiva genericità delle categorie, insufficienti a creare basi solide per l'apprendimento, e, all'opposto, una speculare minuzia rappresentativa, troppo rigida per consentire l'adattamento alla multiforme realtà testuale.

Il raggiungimento di un equilibrio che bilanci le due tendenze opposte, realizzi in modo soddisfacente sia l'obiettivo prescrittivo sia quello descrittivo, e al contempo mantenga una adeguata correttezza sotto il profilo giuridico, è sicuramente una delle sfide più rilevanti in progetti come quello di Adele. Il fatto stesso che si sia trattato di una prima esperienza ha condotto ad un risultato finale sia largamente perfettibile sul piano applicativo. Ancora una volta, però, il lavoro svolto ha prodotto un apporto metodologico che potrà consentire, nel futuro, di approcciare l'attività di elaborazione delle linee guida per la marcatura in maniera meno inconsapevole e, sperabilmente, più efficace.

4. *I limiti giuridici: in particolare il problema dell'(in)spiegabilità.*

Un'analisi tecnica dell'esperienza applicativa che Adele rappresenta potrebbe lasciare intendere, ancora una volta, che i limiti siano solo informatici: la paventata scomparsa del giudice umano ad esclusivo favore di quello robotico, la resurrezione dei dinosauri e l'avvento di Hal 9000 ci attendono dietro l'angolo buio del PGD.

Eppure, nonostante le magnifiche sorti e progressive della digitalizzazione spinta, permane l'argine più antico e ben costruito: quello del diritto. Molte intelligenze artificiali, e sicuramente tutte le intelligenze artificiali basate sul linguaggio naturale condividono un carattere: sono quelle che vengono definite scatole nere (67). Come gli antichi oracoli offrono una risposta, più o meno esplicita, ad una domanda, più o meno criptica, ma non vi accompagnano alcuna spiegazione; quantomeno nessuna spiegazione utile al giurista, o a lui comprensibile. Perché possiamo considerare pacificamente irricevibile come spiegazione il riferimento ad un determinato algoritmo (proprietario, quindi blindato) o ai dati statistici che esso considera come evidenza.

Le magiche scatole nere possono quindi essere considerate ammissibili quando utilizzate per semplificare la vita quotidiana, non certo quando si tratta di addivenire a decisioni che, in misura maggiore o minore, hanno a che fare con l'amministrazione della giustizia. Attualmente, infatti, molti degli algoritmi predittivi non basati su sistemi logici, quindi non coadiuvati dallo sviluppo di sistemi esperti, non offrono alcuna spiegazione rispetto al percorso logico-giuridico che ha consentito loro di addivenire a quella decisione. Ciò per una ragione molto semplice, quasi banale, cioè il fatto che quel

---

(67) P. HALL, N. GILL, *An introduction to machine learning interpretability*, second edition Sebastopol, 2019.

percorso logico-giuridico non esiste. Esiste un'analisi di prossimità, di probabilità statistica, di vicinanza lessicale, ma nessuna spiegazione giuridicamente apprezzabile della soluzione (68).

Ciò significa che, allo stato dell'arte, il famigerato giudice *robot* si porrebbe in sistematico contrasto con l'art. 111 della Costituzione (69). Basterebbe questa considerazione per sciogliere ogni residuo dubbio sulla sostituibilità del giudice con il *cyborg*.

E ciò non solo con riferimento all'utilizzo di tecnologie predittive che si limitano a indicare la percentuale di probabilità che un giudizio si concluda con un certo esito (ad esempio accoglimento o rigetto del singolo motivo), bensì anche immaginando la commistione con le tecnologie quali quelle utilizzate, ad esempio da ChatGPT. Si intende che, anche nell'ipotesi in cui un sistema predittivo predicesse l'esito di una sentenza, il giudice chiedesse poi a un *large language model* di "motivare" quell'esito e il *chatbot* fornisse un testo simile a una "motivazione", non si rispetterebbe comunque l'obbligo di motivazione.

È necessaria, prima di procedere, una premessa: l'ipotesi di cui si tratta nelle prossime righe non è riconducibile a una tecnologia realmente esistente, giacché gli esperimenti di *outcome prediction* condotti all'interno del progetto Adele riguardano esclusivamente la previsione del risultato di sentenze già scritte, i cui i dati di *input* sono le richieste delle parti formulate dal giudice (v. para. 2.2.). Si tratta, quindi, di uno

---

(68) Sul tema S. MULEO, *Redirecting Prodigit: le inutili tentazioni di una sentenza precompilata ed il preferibile indirizzamento dell'IA verso l'obiettivo di una giustizia migliore e più efficiente*, in *Riv. dir. trib.*, 2022, fascicolo monografico, 67; sulla trasformazione da motivazione dell'atto a motivazione rispetto all'esito dell'interrogazione informatica fa riferimento a V. MASTROIACOVO, *Prevedibilità, predittività e umanità del giudicare in materia tributaria*, in *Riv. dir. trib.*, supplemento online, 14 febbraio 2023, 11.

(69) C. RASIA, *La motivazione elastica nel recente caleidoscopio normativo e giurisprudenziale*, in *Riv. trim. dir. proc. civ.*, 2018, 1, 256.

scenario solo astrattamente possibile, in cui la predizione del risultato-sentenza, avvenga sulla base della descrizione dei fatti di causa resa dal soggetto che interroga il sistema e non sia ritratta dal testo di una pronuncia esistente.

Con questa avvertenza, però, si immagini d'unire alla richiesta di predizione di un risultato anche quella di produrre una motivazione di detto risultato, richiesta formulata a ChatGPT, o ad un altro qualsivoglia sistema di *Large Language Modeling* non vietato dal Garante. È vero che noi non sappiamo esattamente come questi sistemi funzionino (sono anch'essi scatole nere), però sappiamo, questo sì, che si basano sul concetto di probabilità statistica del verificarsi di una certa sequenza di parole, non sull'applicazione di regole, siano esse logiche o giuridiche. Ciò significa che, anche nell'ipotesi in cui una motivazione apparisse coerente con quello che, ragionevolmente, avrebbe potuto scrivere un giudice umano, questa sarebbe inammissibile, all'interno del sistema giuridico. Perché l'apparente ragionevolezza estrinseca sarebbe il frutto di un caso, della circostanza statisticamente fortuita dovuta alla vicinanza di parole che in diritto hanno un senso compiuto. Detto in altri termini, la motivazione elaborata da un dispositivo così addestrato non sarebbe né verificabile né falsificabile e per ciò stesso irricevibile.

Se il problema si pone in maniera del tutto insuperabile con riferimento alla sostituzione della funzione giudiziaria, è altrettanto critico se si riporta il tema alla verifica della sostenibilità argomentativa delle decisioni dei giudici di merito attraverso *software* di *argument mining*.

Se, com'è nel caso di Adele, il ritrovamento delle argomentazioni, e da ultimo della struttura logica che vertebrata una sentenza, viene effettuato solo sulla base della ricorrenza linguistica e sull'identificazione di *pattern* lessicali, si rischia di andare incontro ad un diniego di giustizia apoditticamente basato sulle capacità di scrittura del giudice. È evidente che, tutte le volte in cui l'uso dell'IA deve avere un effetto proces-

suale o procedimentale diretto, diverso dal mero *divertissement* per giuristi, l'uso di tecniche basate unicamente sul testo puro sia difficilmente ammissibile e potenzialmente in contrasto coi principi costituzionali, prima ancora che con le norme di diritto positivo.

È pacifico, invece, che il discorso è completamente diverso se si parla di uso dell'IA per attività di mero supporto e ricerca (v. para. 6). Quale che sia la tecnica informatica scelta per sviluppare strumenti di ricerca, organizzazione e classificazione delle sentenze, si tratta di usi che non coinvolgono direttamente il processo decisionale del giudice, né l'attività di indagine dell'Amministrazione Finanziaria, che rimane però estranea al campo di indagine del progetto di Adele. Oltretutto, strumenti più o meno raffinati di automazione sono già presenti nelle banche dati proprietarie che vengono normalmente utilizzate in ambito tributario (e giuridico, più in generale).

##### 5. *Spunti metodologici.*

La costruzione di modelli computazionali in ambito giuridico richiede di affrontare le seguenti questioni fondamentali: quale dominio giuridico modellare e per quale caso d'uso, come acquisire la conoscenza giuridica, come rappresentare la conoscenza giuridica richiesta, quali metodi di inferenza implementare e se il programma "apprenderà".

Le prime domande riguardano la determinazione di quali aree del diritto e quali problemi dovranno essere modellati, nonché del caso d'uso cui il modello sarà applicato. Da quanto detto in relazione alle criticità applicative emerge l'importanza di una chiara progettazione del lavoro alla luce degli obiettivi che si intendono perseguire: alla progettazione dovrebbe essere dedicata una quantità significativa di tempo dal momento che essa orienta tutte le fasi successive.

In tale fase, il lavoro congiunto di giuristi esperti del dominio e informatici è fondamentale. Infatti, affinché la



capacità computazionale del sistema non sia dispersa, il giurista deve avere chiare le possibilità effettivamente perseguibili allo stato dell'arte dell'evoluzione tecnologica, mentre l'informatico deve avere chiari gli aspetti del dominio di riferimento che possono incidere sullo sviluppo del lavoro. Entrambe le figure, comunque, debbono creare il *dataset* pensando agli utenti (a chi è utile il sistema) e agli obiettivi (per cosa è utile il sistema). In questo settore, dunque, può essere particolarmente proficua la mappatura delle esigenze e il confronto preliminare con i soggetti destinatari.

In particolare, è importante sapere se il sistema ha obiettivi puramente scientifici (ad esempio, verificare la fattibilità tecnica di una certa funzionalità) oppure applicativi e, nel secondo caso, se lo strumento è indirizzato al giudice oppure alle parti e se ha funzione di supporto oppure di sostituzione del soggetto deputato allo svolgimento di un certo compito. Infatti, qualora si intenda sviluppare uno strumento che sia suscettibile di applicazioni pratiche, occorre che siano innanzitutto identificate le funzionalità che saranno utili a quel soggetto per perseguire quell'obiettivo. Ad esempio, se si ha intenzione di sviluppare uno strumento a supporto del giudice, può essere utile creare sistemi di ricerca intelligenti per sentenze rilevanti o per la legislazione pertinente. Se si intende invece sviluppare un sistema che fornisca risposte al contribuente, può essere utile sviluppare funzionalità più complesse e lavorare sulla spiegabilità delle risposte fornite dal sistema (70). Inoltre, come si è in parte già detto (v. il para. 4.), occorre interrogarsi sulla compatibilità di tali applicazioni con le disposizioni normative e i principi costituzionali in vigore nell'ordinamento giuridico di riferimento.

L'individuazione delle funzionalità alla luce degli obiettivi e dei soggetti orienta la creazione del *dataset*. Ad esempio, se

---

(70) K. BRANTING, *Semi-supervised Methods for Explainable Legal Prediction*, cit., 214.

l'obiettivo che si persegue ha l'ambizione di svolgere predizioni dell'esito delle domande di parte che siano effettivamente utili, si può immaginare che l'allenamento del sistema debba avvenire anche sugli atti delle parti oltre che sulla sentenza già scritta. Se l'obiettivo è invece quello di estrarre i fattori utilizzati dal giudice in sentenze già scritte, che si assume possano essere utilizzati anche per decidere casi simili in futuro, può essere significativo analizzare la correlazione tra domande e decisioni in sentenze già redatte dal giudice.

Allo stesso modo, se si intende sviluppare un'ontologia che funga da riferimento per gli operatori del diritto, siano essi avvocati, giudici, studiosi o membri dell'Amministrazione Finanziaria, occorre sviluppare un'ontologia che copra i più importanti, se non tutti, gli aspetti dell'imposta. Se si intende verificare la fattibilità del collegamento automatico tra concetti di un'ontologia e un gruppo di sentenze, è sufficiente concentrarsi sugli aspetti dell'imposta che sono toccati nel *dataset* di riferimento. Nella prima ipotesi, dunque, sarà necessaria l'attenta analisi della legislazione; nella seconda ipotesi, sarà ulteriormente necessaria un'attività di lettura e analisi del *dataset* di riferimento. Inoltre, se l'ontologia interviene in un settore armonizzato, è fondamentale riflettere se essa sia destinata al giudice nazionale o ai giudici della Corte di giustizia poiché, solo nel primo caso, occorrerà affrontare tutte le difficoltà relative all'implementazione dell'apparato concettuale europeo nell'ordinamento nazionale per mezzo di concetti e definizioni non sempre semanticamente corrispondenti o identici. Qualora, invece, l'ontologia sia destinata alla Corte di giustizia, sarà opportuno dedicare tempo e risorse alla traduzione dei concetti nelle lingue ufficiali dell'Unione Europea.

Nella creazione del *dataset*, oltre a tenere a mente gli obiettivi che si perseguono, l'esperienza maturata sembra suggerire che sia centrale sfuggire alla tentazione di costruire strumenti che siano di ampissima applicazione. In altri termini, sembra velleitario immaginare lo sviluppo di uno strumento di

predizione generalizzato, oppure lo sviluppo di un'ontologia che copra tutti gli aspetti riconducibili a un'imposta se non all'intero settore fiscale. Quando si costruisce un *dataset*, soprattutto nelle ipotesi di apprendimento supervisionato, o che comunque richiedono un significativo intervento umano, parrebbe opportuno evitare di esagerare la varietà di dati in *input* sapendo di poterli elaborare automaticamente (71). A ciò si dovrebbe prestare particolare attenzione soprattutto quando i dati in *input* sono estremamente vari e diversificati.

In realtà, soprattutto per alcune funzionalità, come la predizione oppure l'*argument mining*, si potrebbe valutare di restringere l'ambito di azione a micro-settori con dati più omogenei (72). Ad esempio nell'ipotesi della predizione, la specificità dei casi sottoposti al giudice e l'ampissima varietà delle circostanze fattuali che danno origine ad essi, renderebbero l'addestramento supervisionato impraticabile. Per allenare un sistema in modo soddisfacente a prevedere l'esito delle domande che una parte può sollevare in un settore vasto come quello di un'intera imposta, servirebbe una quantità di lavoro non accettabile in tempi ragionevoli o che comunque potrebbe rendere lo sforzo e i costi non convenienti rispetto agli esiti. In tal senso, qualora si ritenga opportuno utilizzare l'addestramento supervisionato, sembra ragionevole concentrarsi su un aspetto specifico, individuando una quantità di casi in *input* quantitativamente limitata su cui svolgere un lavoro preciso e approfondito, per poi valutare la possibilità di estendere l'applicazione dei sistemi anche ad altri settori e verificarne in un secondo momento la replicabilità integrale o parziale (73).

Sulla selezione dell'area del diritto e del caso d'uso devono incidere valutazioni di opportunità legate all'acquisizione delle

---

(71) *Ibidem*.

(72) K.D. ASHLEY, *A brief History*, cit., 94.

(73) Si esprime riguardo all'opportunità dell'applicazione di metodi di apprendimento semi-supervisionati su un set limitato di dati annotati per poi applicarlo ad un set più ampio di dati che non sono stati annotati automaticamente, per ridurre gli svantaggi legati ai tempi e costi legati all'annotazione manuale K.

informazioni con cui popolare le rappresentazioni della conoscenza (74). La questione della qualità dei dati si declina in modo diverso a seconda dell'obiettivo che si persegue e si pone in termini particolarmente incisivi in relazione alle tecnologie di apprendimento automatico poiché, come si è mostrato, la conoscenza non viene fornita dall'uomo ma è costruita dal sistema sulla base dei dati a disposizione. La qualità dei dati, la possibilità e le modalità in cui ottenerli sono dunque centrali.

Ad esempio, se si intende sviluppare un'ontologia, non si pongono particolari problemi rispetto all'accessibilità dei dati ma essi potrebbero sorgere qualora si intendesse collegare all'ontologia delle sentenze. Se si intende sviluppare un sistema di *machine learning*, è necessario avere disponibilità di una grande quantità di dati in un formato fruibile al sistema, come nel caso del formato XML. Se si intende rendere fruibile l'*output* del progetto è fondamentale che i dati in *input* siano utilizzabili secondo la disciplina *privacy*. Se le tecnologie di *machine learning* utilizzate fanno riferimento al dato testuale, sembrerebbe un vantaggio disporre di sentenze che siano uniformi dal punto di vista della struttura, delle regole redazionali e, nei limiti del possibile, dello stile utilizzato dal giudice.

Inoltre, sarebbe interessante interrogarsi sull'opportunità di integrare i metodi di *machine learning* basati unicamente sul testo con metodi che prevedono l'estrazione di fattori. Sembra potersi assumere che, se l'estrazione di fattori viene utilizzata con finalità di predizione, le sentenze debbano essere debitamente argomentate e fare riferimento a elementi di fatto. Proprio alla luce della centralità di riferimenti chiari alle circostanze fattuali, ci si potrebbe chiedere se sia più adatto svolgere esperimenti sulle sentenze di primo grado o di se-

---

BRANTING, *Semi-supervised Methods for Explainable Legal Prediction*, in *Proceedings of ICAIL 2019*, 17-21 giugno, Montreal, QC, 22-31.

(74) E. MARELLO, *Prodigit come banca dati intelligente*, in *Riv. dir. trib.*, supplemento *online*, 2022, numero monografico, cit., 96.

condo grado (75). Inoltre, se si intende sviluppare un sistema che tenti di prevedere l'esito delle domande di parte, occorre che gli atti di parte siano accessibili. Diversamente, se si vuole sviluppare un sistema esperto, la qualità dei dati in *input* ha un rilievo meno dirimente, ma è necessario arruolare (più) persone con competenze giuridiche per compilare le regole con cui il sistema può analizzare gli scenari problematici e ricordare che tali regole dovranno essere modificate, con lo stesso metodo, al mutare della legislazione o della giurisprudenza.

È evidente che lo sviluppo di sistemi di IA è favorito dalla digitalizzazione e dalla fruibilità dei dati. In altri termini in un settore, come quello della giustizia tributaria, che astrattamente si presterebbe all'applicazione delle tecnologie perché i dati sono tanti, ma in cui essi sono di qualità scadente o utilizzabile solo a fronte di un significativo lavoro umano, l'impiego delle tecnologie è sfavorito e sembra dover passare dalla possibilità che in una fase antecedente i dati siano resi accessibili in formato utilizzabile. In tal senso, l'iniziativa di esortare i giudici a redigere le sentenze in un formato leggibile, come il formato PGD, può essere un utile punto di partenza ma essa sarebbe potenziata se vi si accompagnasse l'inclusione di ulteriori metadati. Anche in tal caso, comunque, tutte le attività esaminate (*outcome prediction*, *argument mining*, *network analysis* e ontologia formale) sarebbero complicate dalla commistione tra aspetti relativi alla disciplina sostanziale e procedurale nelle sentenze in materia fiscale.

Inoltre, può essere utile soffermarsi sull'ordine delle attività svolte. L'esperienza applicativa maturata dimostra che, nel caso in cui si intenda svolgere più attività diverse rispetto al medesimo *dataset*, ci sono attività che ha senso svolgere prima di altre. Ad

---

(75) Tuttavia, secondo L.K. BRANTING, *Scalable and explainable legal prediction*, cit., 217 le sentenze sia di primo grado sia di appello non sono generalmente adatte all'approccio di rappresentazione dei fatti e delle argomentazioni secondo categorie delle decisioni, a causa dell'ampia varietà di fatti che danno origine alle cause e di tipi di decisioni.

esempio, nel nostro caso, lo sviluppo di un'ontologia consente di creare una tassonomia utile potenzialmente anche per le attività di *machine learning*. Perciò sembrerebbe ragionevole che tale attività sia cronologicamente anteriore o contemporanea alle attività relative alla predizione e all'*argument mining*. Infatti, da un lato, l'ontologia potrebbe garantire agli esperti del settore occasione di sistematizzazione della conoscenza su cui si troveranno a lavorare in seguito; dall'altro lato, essa potrebbe integrare le attività di analisi sul testo puro per valutarne l'impatto sugli esiti ottenuti in relazione alle singole funzionalità.

Infine, si vuole sottolineare che alcuni degli aspetti appena menzionati, se i suddetti strumenti fossero effettivamente usati per assistere le attività giudiziarie, potrebbero essere resi obbligatori qualora la proposta di Regolamento sull'intelligenza artificiale (76). Infatti, l'art. 10 c. 2 della Proposta prevede che i *set* di addestramento dei sistemi di IA ad alto rischio siano sottoposti a pratiche di *governance* che riguardino scelte progettuali pertinenti, raccolta dei dati, operazioni di annotazione, l'individuazione di lacune o carenze nei dati ecc. (77)

#### 6. *Predizione e non perversione: l'uso dell'IA a supporto e non in sostituzione del giudice.*

Dopo due anni di lavoro nel piccolo cabotaggio dell'intelligenza artificiale applicata alle Commissioni Tributarie, sentir

---

(76) Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione, Bruxelles, 21.4.2021, COM(2021) 206 final. Occorre specificare che il Parlamento europeo in data 14 giugno 2023 ha dato il via libera AI Act.

(77) L'Allegato III prevede che i sistemi di IA ad alto rischio a norma dell'articolo 6, paragrafo 2, sono anche i sistemi di IA destinati ad assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti e, dunque, sebbene ciò richiederebbe un'analisi più approfondita che esula dagli obiettivi del presente contributo, tali requisiti potrebbero divenire obbligatori qualora i sistemi fossero effettivamente implementati.

preconizzare l'imminente arrivo del giudice robot produce un effetto di straniamento. Come Montale, in questo periodo di tempo il *machine learning* ci ha piuttosto detto ciò che non siamo e ciò che non vogliamo, ma soprattutto ciò che non possiamo (fare). Sicuramente il *machine learning* non può, allo stato dell'arte, predire il futuro o decidere di un caso, per la banale ragione per cui l'intelligenza digitale difetta di una capacità essenziale allo scopo che è quella di ragionamento.

Si potrebbe forse immaginare un *software* che sia in grado di redigere una sentenza sulla base di precisi *input* offerti dall'uomo, quali il riassunto del fatto, delle questioni sottoposte, la qualificazione della fattispecie e l'esito della decisione. Ma, anche in questa ipotesi, la responsabilità della decisione, l'onere dello studio del fascicolo, l'applicazione del ragionamento giuridico alla soluzione del caso spetterebbe al giudice tradizionale, quello umano. Se ne ricaverebbe un risparmio di tempo? Forse, ma non per certo, considerato che la piena assunzione di responsabilità dell'uomo lo obbligherebbe, anche in questo scenario futuribile, a rivedere, rimaneggiare e verificare dettagliatamente il testo prodotto dal suo *ghost writer* digitale. Oltretutto, l'obiettivo di far scrivere all'intelligenza artificiale esattamente ciò che si vuole in termini di contenuti obbliga il questuante ad una formulazione precisa fin oltre il limite della pedanteria della questione sottoposta. Con la conseguenza che un tale rigore redazionale verrebbe richiesto al giudice in ognuno dei passaggi descrittivi da sottoporre alla macchina perché questa possa produrre la sentenza. Quasi da far sembrare più semplice il ritorno a metodi antichi, vale a dire alla produzione in proprio della sentenza stessa.

Ci si può chiedere se queste perplessità possano essere superate dalla scelta di limitare l'applicazione dell'IA a temi circoscritti in settori estremamente specifici (78). È indubita-

---

(78) Si sono espressi, tra gli altri, riguardo ai settori in cui sarebbe opportuno utilizzare l'intelligenza artificiale con finalità predittive e, in particolare, sulla

bile che una scelta di questo genere potrebbe rendere più efficace l'applicazione di strumenti predittivi o generativi, perché semplificherebbe la preparazione del *training set* e al contempo consentirebbe di costruirlo in maniera più efficace e, per l'effetto, condurrebbe probabilmente a risultati migliori e più soddisfacenti. Non è, però, ugualmente indubbio che optare per questa strada, cioè per la totale automazione delle decisioni in un microsettore, sia destinata ad essere la scelta più efficace ed economica. Come si è detto addestrare un sistema di *machine learning* con tecniche di addestramento supervisionato, soprattutto se puramente basato su analisi linguistica, è un'attività lunga e complessa, in cui l'impegno umano richiesto potrebbe talvolta risultare sproporzionato rispetto all'obiettivo da raggiungere.

Si potrebbe forse immaginare il ricorso a tecnologie più antiche dal punto di vista informatico, quali quelle connesse all'uso di sistemi esperti, che consentirebbero di offrire una risposta certa, perché basata su di una consolidata base di regole, su questioni puntuali, come ad esempio quelle relative al calcolo dei termini decadenziali, una volta risolte eventuali questioni interpretative. È però chiaro che una soluzione del genere, rappresenterebbe più che una proposta di sostituzione del soggetto giudicante uno strumento, seppur raffinato, posto a suo supporto e che, in ogni caso, dovrebbe essere continuamente aggiornato.

All'irrealizzabilità tecnologica di una predizione che si propone di sostituire il giudice si affiancano le questioni giuridiche in parte già anticipate (v. para. 4) e ampiamente

---

necessità di avere cause seriali a bassa complessità, con presupposti fattuali semplici, poche norme da applicarsi e contenuto immediato F. FARRI, *La giustizia predittiva*, cit. e A. CARINCI, *Il processo tributario alla prova della giustizia predittiva*, in *Riv. dir. trib.*, 2022, fascicolo monografico, cit., 27.



analizzate nella dottrina sia tributaria (79), sia informatico-giuridica (80).

La prima ha sottolineato, tra gli altri aspetti, che profili particolarmente problematici sarebbero legati alla circostanza che nel giudizio tributario spesso il giudice procede alla ride-terminazione quantitativa della pretesa e delle sanzioni (compito che l'IA difficilmente potrebbe svolgere) (81), all'estrema eterogeneità, per forma e complessità, di redazione delle sentenze (82) e al rischio di limitare l'emersione di nuovi orientamenti giurisprudenziali (83).

La seconda ha evidenziato che i rischi legati all'utilizzo delle tecnologie di IA in chiave sostitutiva del giudice sono legati al fatto che il sistema non operi il *distinguishing* nelle situazioni in cui un essere umano lo farebbe (84), al cd. *effetto gregge*, cioè alla situazione in cui il giudice faccia propria la proposta dell'algoritmo senza assumersi la responsabilità della decisione con la conseguente cristallizzazione della giurisprudenza (85).

In particolare, l'assenza di adeguata spiegazione giuridica e la centralità della similarità rispetto ai casi precedenti, senza tenere in considerazione le maggiori conoscenze accessibili al giudice (86), non rendono plausibile né auspicabile l'utilizzo

---

(79) Per una ricognizione degli interventi più recenti si veda nota 60.

(80) Ne propongo una ricognizione A. SANTOSUOSSO, G. SARTOR, *La giustizia predittiva*, cit.

(81) A. CARINCI, *Il processo tributario*, cit., 27.

(82) *Ivi*, 28.

(83) A. MARCHESELLI, *Uomo, automazione, giustizia: le relazioni pericolose e la morte della jusdiversità*, in *Riv. dir. trib.*, 2022, numero monografico, 9.

(84) C. O'NEIL, *Weapons of math destruction: how big data increases inequality and threatens democracy*, in *Crown business*, 2016 sottolinea che i critici delle decisioni automatizzate affermano che le decisioni non possono mai essere realmente individualizzate.

(85) A. SANTOSUOSSO, G. SARTOR, *La giustizia predittiva*, cit.

(86) F. BEX, H. PRAKKEN, *The legal prediction industry*, cit. Sul punto si veda anche il documento della Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) approvato il 3-4 dicembre 2018 a Strasburgo.

dell'intelligenza artificiale in chiave sostitutiva e generalizzata (87).

L'esperienza applicativa, quindi, si unisce alla critica sistemica ed entrambe sembrerebbero suggerire per il futuro di prestare una maggiore attenzione non tanto alla all'utilizzo dell'IA in chiave sostitutiva del giudice (l'IA *in sostituzione*), ma anche e soprattutto all'analisi delle potenzialità e dei rischi delle tecnologie di IA e dei sistemi esperti a supporto del giudice (l'IA *a supporto*) (88).

L'esperienza del progetto Adele non comprende la modellazione di regole per dei sistemi esperti, di conseguenza sarebbe ultroneo, ma soprattutto privo di riscontri sperimentali, addentrarsi in suggestioni relative all'uso di queste tecnologie. Si torna però a ribadire che l'integrazione di tali sistemi all'interno di esperimenti di cui abbiamo avuto una diretta esperienza avrebbe forse consentito di ottenere risultati più soddisfacenti anche sotto il profilo giuridico, non solo informatico. Ad esempio, nel caso dell'*argument mining*, potrebbe concorrere allo sviluppo di strumenti di supporto utili a tutte le parti e non solo al giudice, tesi alla verifica di eventuali aporie logiche nelle motivazioni delle sentenze, tali da giustificare una censura delle parti in sede di giudizio di appello o di Cassazione.

Al contempo, giunti al termine di Adele, il maggior potenziale di sviluppo per un uso dell'IA che moltiplichi le potenzialità di ricerca e organizzazione della conoscenza, non solo a favore dei giudici, ma di tutti gli operatori del diritto, sembra venire dallo sviluppo di ontologie e *citation network analysis*.

Le ontologie potrebbero essere utilizzate per l'organizzazione delle sentenze secondo tassonomie strutturate e più precise di quelle attualmente in uso, che aiutino la ricerca nelle

---

(87) I sistemi predittivi potrebbero eventualmente essere valutati per indicare la decisione probabile del giudice medio ai contribuenti e ai loro difensori.

(88) Si veda anche S. DORIGO, *Intelligenza artificiale*, cit.

banche dati a scopo di lavoro e studio; per la creazione di mappe concettuali intelligenti che consentano di visualizzare e accedere in maniera semplice a temi connessi ma distinti in una materia stratificata e con cospicui rinvii ad altri settori del diritto come quella fiscale; per l'estrazione di parole chiave in sentenze particolarmente omogenee dal punto di vista lessicale (89); come strumento di navigazione e di visualizzazione in banche dati intelligenti (90).

Lo sviluppo di un'ontologia richiederebbe un ampio lavoro di ricostruzione dell'apparato concettuale da parte degli esperti del settore e dovrebbe confrontarsi con tutte le difficoltà che la creazione di qualsiasi tassonomia crea. Si tratta, in questo caso, di rischi che per la maggior parte esisterebbero anche nell'ipotesi di svolgimento umano dell'attività. Le tecnologie, in questo caso, aiuterebbero semplicemente al reperimento più efficiente del materiale e, se possibile (non siamo ancora in grado di dirlo), all'eventuale classificazione automatica dei dati sulla base della tassonomia (91). Nel caso in cui l'ontologia, creata da un esperto del settore, sia utilizzata come strumento di mera navigazione in una banca dati, non si pongono rischi diversi dall'ipotesi in cui tale strumento non

---

(89) Sulle ragioni, anche non giuridiche, dell'uniformità lessicale e sintattica della Corte di giustizia si vedano L.N. BROWN, T. JACOBS, *The Court of Justice of the European Communities*, Londra, 2000, 23 ss. e K. McAULIFFE, *Hybrid Texts and Uniform Law? The multilingual case law of the Court of Justice of the European Union*, in *International Journal for the Semiotics of Law*, 2011, vol. 24, n. 1, 97 ss.

(90) E. MARELLO, *Prodigit come banca dati intelligente*, cit., 94 suggerisce che nella costruzione di una potenziale futura banca dati per la giustizia tributaria ci sia lo spazio per un ampio uso dell'intelligenza artificiale in modo molto più affidabile, costruttivo e utile di quanto possa avvenire nell'oracolare campo della predizione.

(91) Sulla centralità della classificazione, il problema è forse limitabile consentendo lo scaricamento a pacchetti indicato sopra, magari accompagnato da più indicizzazioni del materiale, orientate a diversi criteri (e non da una sola indicizzazione di rilevanza). Poiché le classificazioni sono un esercizio tipico da intelligenza artificiale, ecco un altro uso utile che si potrebbe dare alle risorse del progetto: proporre la possibilità di definire in maniera personalizzabile i criteri di rilevanza (e quindi l'*output* dell'indicizzazione).

esistesse e la ricerca avvenisse con le modalità tradizionali (ad esempio, con la ricerca a testo libero) (92). Nell'ipotesi in cui, invece, l'ontologia sia utilizzata per la classificazione automatica dei dati, si porrebbero i rischi relativi all'erronea classificazione del materiale e, quindi, alla possibilità di reperimento del materiale.

Per quanto riguarda la *citation network analysis*, se fossero disponibili documenti nazionali, vale a dire sentenze tributarie di merito, sufficientemente puliti ed elaborati (93), sul modello della giurisprudenza europea, molto potrebbe essere fatto. Innanzitutto, un'organizzazione di rete delle sentenze, sulla base della riconducibilità ai medesimi riferimenti legislativi o giurisprudenziali produrrebbe l'effetto, sempre auspicato, di una maggiore conoscibilità della giurisprudenza tributaria di merito. La *network analysis*, infatti, non nasce con l'obiettivo di predire il futuro bensì con quello, più modesto ma forse più utile, di aiutare a comprendere il passato.

Si immagini di poter sapere immediatamente, anche con il supporto di una rappresentazione grafica, a quali sentenze di Cassazione il giudice di merito sta facendo riferimento, magari a fronte di un tema dibattuto all'interno della stessa giurisprudenza di legittimità, e a loro volta, quali siano le pronunce di merito che, su base nazionale, contengono i medesimi riferimenti. Non solo, si immagini anche di poter verificare e comparare quali siano le altre posizioni interpretative presenti sul tema e quando, cronologicamente, le due soluzioni opposte sono state prospettate dai supremi giudici.

Nessuna di queste ipotesi contiene un vincolo per il giudice, non nella forma di un suggerimento sulla decisione da

---

(92) Si sta parlando dell'ipotesi in cui, ad esempio, sia il giudice ad individuare i concetti dell'ontologia cui la sentenza è legata e, dunque, l'ontologia stessa è utilizzata come strumento di visualizzazione e navigazione.

(93) Certo, questa prospettiva richiederebbe, come dice A. CARINCI, *Quale giustizia predittiva?*, cit., 29, l'avvento di giudici più disciplinati e attenti a produrre sentenze di cui venga assicurata una pronta lettura e comprensibilità.

prendere, né in quella dell'introduzione di precedenti rappresentati in maniera sempre più stringente, tanto da diventare vincolante. Al contempo, si tratta di strumenti che permettono di ampliare la conoscenza di un materiale vasto e spesso disorganizzato qual è la giurisprudenza pregressa. Una conoscenza che, auspicabilmente, dovrebbe poter concedere non solo ai giudici, ma anche agli avvocati e agli studiosi, una maggiore consapevolezza, una conseguente capacità di decostruzione dei percorsi giuridico-interpretativi consolidati e, in sintesi, una maggiore libertà.

#### 7. *Considerazioni conclusive.*

All'esito delle molte considerazioni svolte, ma soprattutto dell'esperienza applicativa diretta, si può forse concludere che, nel mondo del diritto, l'intelligenza artificiale corre molto meno veloce di quanto il suo ufficio marketing voglia far credere.

Non solo perché i vincoli giuridici non sono orpelli astratti, ma fondamentali presidi di giustizia e democraticità del sistema, sebbene questo rappresenti un limite non facile da superare, ma anche per la presenza di sfide metodologiche e applicative di non semplice soluzione.

La prima sfida è quella del tempo. La messa a terra del *machine learning* in qualunque ambito, non solo in quello giuridico, richiede un significativo investimento di tempo ed energie da parte di tutti i soggetti coinvolti nella sua realizzazione. L'impegno riguarda innanzitutto le attività prodromiche, forse quelle più scientificamente rilevanti: la definizione delle linee guida per il *tagging*, la selezione del dominio e dei documenti da inserire nel *dataset*, la preparazione della tassonomia su cui si fonda l'ontologia, la corretta individuazione dell'ipotesi di lavoro e dei risultati attesi. Una volta completata questa fase deve iniziare l'attività di annotazione vera e propria, che è forse l'impegno che assorbe più tempo con un

minore riscontro scientifico, perché non si limita alla prima marcatura del testo, ma comporta anche la revisione e la modifica per ogni, inevitabile, modifica del *set* di *tag* (94). A ciò si aggiungono gli esperimenti condotti dagli esperti informatici, nonché tutti quei passaggi di “aggiustamento” necessari al raggiungimento del risultato finale. La variabile “tempo”, perciò, dovrebbe essere sempre presa in considerazione per chiedersi se “ne valga la pena”, per capire se il risultato atteso giustifica un tale impegno di risorse di personale altamente qualificato.

La seconda sfida è quella delle fonti. L'esperienza applicativa di Adele ha reso evidente come la selezione delle fonti documentali sia fondamentale per ottenere un risultato sia qualitativamente sia metodologicamente significativo. Le fonti sono dirimenti per orientare il lavoro sotto molteplici aspetti, come già evidenziato nel testo: la loro reperibilità, qualità, formato e organizzazione. Il processo di creazione del *dataset* dipende, nella sua ricchezza e organicità, dalla presenza di banche dati pubbliche e accessibili, in cui i documenti presenti siano organizzati in maniera razionale, così da evitare dispersioni di contenuto e, anche, ridurre l'impegno umano per il loro reperimento. Dati di qualità, nel senso di documenti che siano originariamente disponibili in un formato *machine readable*, minimizzano il rischio di errori frutto di errori testuali sul piano simbolico, e rendono quindi ogni risultato più attendibile. Documenti organizzati, con strutture formali riconoscibili e, auspicabilmente, già dotati di un set di marcatura “validato” dall'organo proprietario consentono di velocizzare

---

(94) È, peraltro, opportuno ricordare che costo dell'addestramento è un sacrificio scientifico per gli esperti del dominio in piccoli esperimenti come Adele, un costo sociale ed ecologico nell'elaborazione di enormi sistemi di *machine learning*, com'è il caso di Open AI. Si veda, tra gli altri, il recente articolo di A. CASILLI, *Il lato oscuro dell'algoritmo è la forza lavoro*, <https://ilmanifesto.it/chatgpt-il-lato-oscuro-dell'algoritmo-e-la-forza-lavoro-che-addestra-la-macchina-digitale> (ultimo accesso: 26 ottobre 2023).

il lavoro perché riducono la necessità di intervento umano in fase di preparazione e marcatura a tutto favore della valutazione dei risultati e del raffinamento delle questioni di ricerca.

L'ultima sfida è quella degli obiettivi. Proprio perché la ricerca nel settore della *Legal Analytics* richiede un'applicazione così minuziosa e metodologicamente complessa, la preoccupazione maggiore dev'essere quella relativa alla corretta individuazione del campo di ricerca e delle aspettative che vi si ripongono. L'esperienza applicativa ha dimostrato come l'approccio a settori troppo ampi e il sovrapporsi di troppe ambizioni rendano il lavoro più complesso e ne aumentino il margine d'errore, col rischio di ridurre specularmente la sua utilità. L'auspicio per il futuro, quindi, è quello di saper selezionare oculatamente quei temi che, nel diritto tributario, siano più adatti all'applicazione delle tecnologie dell'intelligenza artificiale e rispetto cui sia possibile immaginare e sviluppare funzionalità realistiche e utili a moltiplicare e migliorare gli strumenti di cui ogni operatore del diritto può disporre per lo studio e l'applicazione della materia.

