

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Deceitful media: Artificial intelligence and social life after the Turing Test

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1768312> since 2024-10-29T13:52:59Z

Publisher:

Oxford University Press

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

From: Simone Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test* (New York: Oxford University Press, 2021), 1-17.

Introduction

In May 2018, Google gave a public demonstration of its ongoing project Duplex, an extension of Google Assistant programmed to carry out phone conversations. Google's CEO Sundar Pichai presented the recording of a conversation in which the program mimicked a human voice to book an appointment with a hair salon. Duplex's synthetic voice featured pauses and hesitation in an effort to sound more credible. The strategy appeared to work: the salon representative believed she was speaking with a real person and accepted the reservation.¹

In the following weeks, Google Duplex's apparent achievements attracted praise, but also criticism. Commentaries following the demonstration highlighted two problems about the demo. On the one side, some contended that Duplex operated "straight up, deliberate deception," opening new ethical questions regarding an Artificial Intelligence (AI)'s capacity to trick users into believe it is human.² On the other side, some expressed doubts about the authenticity of the demo. They pointed to a series of oddities in the recorded conversations: the businesses, for instance, never identified themselves, no background noise could be heard, and the reservation-takers never asked Duplex for a contact number. This suggested that Google might have doctored the demo, faking Duplex' capacity to pass as human.³

The controversy surrounding Duplex reflects a well-established dynamic in the public debate about AI. Since its inception in the 1950s, the achievements of AI have often been discussed in binary terms: AI is either attributed exceptional powers or dispelled as a delusion

and a fraud.⁴ Time after time, the gulf between these contradictory assessments jeopardized our capacity to recognize that the true impact of AI is more nuanced and oblique than usually acknowledged. The same risk is present today, as commentators appear to believe that the question should be whether or not Duplex is able to pass as a human. However, even if Google's gadget proved unable to pass as human, we should not believe the illusion to be dispelled. Also in the absence of deliberate misrepresentation, AI technologies entail forms of deception that are perhaps less evident and straightforward, but deeply impact on our societies. We should regard deception not just as a possible way to employ AI, but as a constitutional element of these technologies. It is as central to AI's functioning as the circuits, software, and data that make it run.

This book argues that, since the beginning of the computer age, researchers and developers have explored how users are led to believe that computers are intelligent. Examining the historical trajectory of AI from its origins to the present day, I show that AI scientists incorporated knowledge about users into their efforts to build meaningful and effective interactions between humans and machines. I call, therefore, for a recalibration of the relationship between deception and AI that critically questions how computing technologies draw on specific aspects of users' perception and psychology in order to create "AI."

One of the foundational texts for AI research, Alan Turing's *Computing Machinery and Intelligence* (1950), set up deception as a likely outcome of interactions between humans and intelligent computers. In his proposal for what is now commonly known as the Turing Test, Turing suggested to evaluate computers based on their capacities to deceive human judges into believing they were human. Although tricking humans was never the main objective of AI, computer scientists adopted Turing's intuition that whenever communication with humans is

involved, the behavior of the human users informs the meaning and impact of AI just as much as the behavior of the machine itself. As new interactive systems that enhanced communications between humans and computers were introduced, AI scientists began more seriously engaging with questions of how humans react to seemingly intelligent machines. The way this dynamic is now embedded in the development of contemporary AI voice assistants such as Google Assistant, Amazon's Alexa and Apple's Siri signals the emergence of a new kind of interfaces, which mobilize deception to manage the interaction between users, computing systems and Internet-based services.

After Turing's field-defining proposal, AI has coalesced into a disciplinary field within cognitive science and computer science, producing an impressive range of technologies that are now of public use, from machine translation to the processing of natural language, from computer vision to the interpretation of medical images. Researchers in this field nurtured the dream – cherished by some scientists while dismissed as unrealistic by others – of reaching “strong” AI, i.e. a form of machine intelligence that is practically indistinguishable from human intelligence. Yet, while debates have largely focused on the possibility that the pursuit of strong AI will lead to forms of consciousness similar or alternative to human life, where we have landed might be more accurately described as the creation of a range of technologies providing an *illusion* of intelligence – in other words, not the creation of intelligent beings, but of technologies that are perceived by humans as intelligent.

Reflecting broader evolutionary patterns of narratives about technological change, the history of AI and computing has until now been mainly discussed in terms of technological capability.⁵ Even today, the proliferation of new communicative AI systems is mostly explained as a technical innovation sparked by the rise of neural networks and Deep Learning.⁶ While

approaches to the emergence of AI usually emphasize evolution in programming and computing technologies, *Deceitful Media* focuses on how the development of AI also built on knowledge about users.⁷ Taking up this point of view helps realise the extent to which tendencies to project humanity onto things makes AI potentially disruptive for social relationship and everyday life in contemporary societies. This book, therefore, reformulates the debate on AI on the basis of a new assumption: that what machines are changing is primarily us, humans. “Intelligent” machines might one day revolutionize life; they are already transforming how we understand and carry out social interactions.

Since AI’s emergence as a new field of research, many of its leading researchers have professed to believe that humans are fundamentally similar to machines and, consequently, that it is possible to create a computer that equals or surpasses human intelligence in all aspects and areas. Yet entertaining a similar tenet does not forcefully contrast, but is often complementary to, the idea that existing AI systems provide only the illusion of human intelligence. Throughout the history of AI, many acknowledged the limitations of present systems, and focused their efforts on designing programs that provide at least the appearance of intelligence; in their view, ‘real’ or ‘strong’ AI would come through further progress, with their own simulation systems representing just a step in that direction.⁸ Understanding how humans engage in social exchanges, and how they can be led to treat things as social agents, became instrumental to overcoming the limitations of AI technologies. AI researchers established a direction of research based on the design of technologies that cleverly exploit human perception and expectations to give users the impression of employing or interacting with intelligent systems. While such tradition has not yet been studied as such, this book demonstrates that looking at its development across time is essential to understand contemporary AI systems programmed to engage socially

with humans. In order to pursue this agenda, however, the problem of deception and AI should be formulated under new terms.

On humans, machines and ‘banal deception’

When the great art historian Ernst Gombrich started his inquiry into the role of illusion in the history of art, he realized that figurative arts emerge within an interplay between the limits of tradition and the limits of perception. Artists have always incorporated deception in their work, drawing on their knowledge of both convention and mechanisms of perception to achieve certain effects on the viewer.⁹ But who would blame a gifted painter for employing deceit by playing with perspective or depth to make a tableau look more convincing and ‘real’ in the eyes of the observer?

While this is easily accepted from an artist, the idea that a software developer employs knowledge about how users are deceived in order to improve human-computer interaction is likely to encounter concern and criticism. In fact, because the term is usually associated with malicious endeavors, the AI and computer science communities have proven resistant to discussing their work in terms of deception, or have discussed deception as an unwanted outcome.¹⁰ This book, however, contends that deception is a constitutive element of human-computer interactions rooted in AI technologies. We are, so to say, programmed to be deceived, and modern media have emerged within the spaces opened by the limits and affordances of our capacity to fall into illusion. Despite their resistance to consider deception as such, computer scientists have worked since the early history of their field to exploit the limits and affordances of our perception and intellect.¹¹

A wealth of research in areas such as social psychology, philosophy and sociology has shown that deception and self-deception are inescapable facts of social life with a functional role in social interaction and communication.¹² Although situations in which deception is intentional and manifest, such as frauds, scams and blatant lies, shape popular understandings of deception, scholars have underlined the more disguised, ordinary presence of deception in everyday experience.¹³ Many forms of deception are not so clear-cut, and in many cases, deception is not even understood as such.¹⁴

Moving from a phenomenological perspective, philosopher Mark A. Wrathall influentially argued that our capacity to be deceived is an inherent quality of our experience. While deception is commonly understood in binary terms, positing that one might either be or not be deceived, Wrathall contends that such dichotomy does not account for how people perceive and understand external reality: “it rarely makes sense to say that I perceived either truly or falsely” since the possibility of deception is ingrained in the mechanisms of our perception. If, for instance, I am walking in the woods and believe to see a deer on my side where in fact there is just a bush, I am deceived; yet the same mechanism that made me see a deer where it wasn’t – i.e. our tendency and ability to identify patterns in visual information – would have helped me, in another occasion, to identify a potential danger. The fact that our senses have fallacies, Wrathall points out, represents a limit as much as a resource for human perception and is functional to our ability to navigate the external world.¹⁵ From a similar point of view, cognitive psychologist Donald D. Hoffman recently proposed that evolution has shaped our perceptions into useful illusions that help us navigate the physical world, but can also be manipulated through technology, advertising and design.¹⁶

Indeed, the institutionalization of psychology in the late nineteenth and early twentieth century already signaled the discovery that deception and illusion were integral, physiological aspects in the psychology of perception.¹⁷ Understanding deception was important not much or not only to study how people misunderstand the world, but also to study how they perceive and navigate it.¹⁸ During the nineteenth and twentieth century, the accumulation of knowledge on how people are deceived informed the development of a wide range of media technologies and practices, whose effectiveness exploited the affordances and limitations of our senses of seeing, hearing and touching.¹⁹ As I demonstrate in this book, AI has continued this tradition of technologies that mobilize our liability to deception to produce their outcomes. AI scientists collected information and knowledge on how users react to machines exhibiting the appearance of intelligent behaviors, incorporating this knowledge into the design of software and machines.

One potential objection to this approach is that it dissolves the very concept of deception by equating it with “normal” perception. I contend, however, that refusing a binary understanding of deception helps realize that deception involves a wide spectrum of situations that have very different outcomes but also common characteristics. If on one end of the spectrum there are explicit attempts to mislead, commit fraud and tell lies, on the other end there are forms of deception that are not so clear-cut and that, in many cases, are not understood as such.²⁰ Only by identifying and studying less evident dynamics of deception we can develop a full understanding of more evident and straight-out instances of deception. In pointing to the centrality of deception, therefore, I do not intend to suggest that all forms of AI have hypnotic or manipulating goals. My main goal is not to establish if AI is ‘good’ or ‘bad’, but to explore a crucial dimension of AI and interrogate how we should proceed in response to this.

Home robots such as Jibo or companion chatbots such as Replika, for example, are designed to appear cute and to awaken sentiments of empathy in their owners. This looks in themselves harmless and benevolent: these technologies simply work better if their appearance and behavior stimulate positive feelings in their users.²¹ The same characteristics, however, will appear less innocent if the companies producing these systems start profiting from these feelings in order to influence the political opinions of users. Home robots and companion chatbots, together with a wide range of AI technologies programmed to enter in communication with humans, structurally incorporate forms of deception: elements such as their appearance, their humanlike voice and the use of specific language expressions are designed to produce specific effects in the user. What makes this less or more acceptable is not the question if there is or there isn't deception, but rather the outcomes and the implications of the deceptive effects produced by any given AI technology. Broadening the definition of deception, in this sense, can lead to improving our comprehension of the potential risks of AI and related technologies, counteracting the power of the companies that gain from the user's interactions, and stimulating broader investigations of whether this interaction poses any potential harm to the user.

To distinguish from straight-out and deliberate deception, I propose the concept of *banal deception* to describe deceptive mechanisms and practices that are embedded in media technologies and contribute to their integration into everyday life. Banal deception entails mundane, everyday situations in which technologies and devices mobilize specific elements of user's perception and psychology – for instance in the case of AI, the all-too-human tendency to attribute agency to things or personality to voices. The word “banal” describes things that are dismissed as ordinary and unimportant; my use of this word aims to underline that these

mechanisms are often taken from granted, despite their significant impact on media's uses and appropriations, but are deeply embedded in our everyday, 'ordinary' life.²²

Different from approaches to deliberate or straight-out deception, banal deception does not understand users and audiences as passive or naïve. On the contrary, audiences actively exploit their own capacity to fall into deception in sophisticated ways – for example, through the entertainment they enjoy when they fall into the illusions offered by cinema or television. The same mechanism resonates with the case of AI. Studies in human-computer interaction consistently show that users interacting with computers apply norms and behaviors that they would adopt with humans, even if these users perfectly understand the difference between computers and humans.²³ At first glance, this seems incongruous, as if users resist and embrace deception simultaneously. The concept of banal deception provides a solution to this apparent contradiction. I argue that the subtle dynamics of banal deception allow users to embrace deception so that they can better incorporate AI into their everyday lives, making AI more meaningful and useful to them. This does not mean that banal deception is harmless or innocuous. Structures of power often reside in mundane, ordinary things, and banal deception may finally bear deeper consequences for our societies than the most manifest and evident attempts to deceive.

Throughout this book, I identify and highlight five key characteristics that distinguish banal deception. The first one is its everyday and ordinary character. When researching people's perceptions of AI voice assistants, Andrea Guzman was surprised by what she sensed was a discontinuity between the usual representations of AI and the responses of her interviewees.²⁴ AI is usually conceived and discussed as extraordinary: a dream or a nightmare that awakes metaphysical questions and challenges the very definition of what means to be human.²⁵ Yet

when Guzman approached users of systems such as Siri, the AI voice assistant embedded in iPhones and other Apple devices, she did not find that they were questioning the boundaries between humans and machines. Instead, participants were reflecting on themes similar to those that also characterize other media technologies. They were asking if using the AI assistant made them lazy, or if it was rude to talk to the phone in the presence of others. As Guzman observes, “neither the technology nor its impact on the self from the perspective of users seemed extraordinary; rather, the self in relation to talking AI seemed, well, ordinary – just like any other technology.”²⁶ This ordinary character of AI is what makes banal deception so imperceptible but at the same time so consequential. It is what prepares AI technologies to be integrated into the fabrics of everyday experience and, as such, into the very core of our identity and self.²⁷

The second characteristic of banal deception is functionality. Banal deception always has some potential value to the user. Human-computer interaction has regularly employed representations and metaphors to build reassuring and easily comprehensible systems, hiding the complexity of the computing system behind the interface.²⁸ As noted by Michael Black, “manipulating user perception of software systems by strategically misrepresenting their internal operations is often key to producing compelling cultural experiences through software.”²⁹ Using the same logic, communicative AI systems mobilize deception to achieve meaningful effects. The fact that users behave socially when engaging with AI voice assistants, for instance, has an array of pragmatic benefits: it makes it easier for users to integrate these tools into domestic environments and everyday lives, and presents possibilities for playful interaction and emotional reward.³⁰ Being deceived, in this context, is to be seen not as a misinterpretation of the user, but as a response to specific affordances coded into the technology itself.

The third characteristic of banal deception is obliviousness: the fact that the deception is not understood as such but taken for granted and unquestioned. The concept of “mindless behavior” has been already used to explain the apparent contradiction, mentioned above, by which AI users understand that machines are not human but still to some extent treat them as such.³¹ Researchers drew from cognitive psychology to describe mindlessness as “an overreliance on categories and distinctions drawn in the past and in which the individual is context-dependent and, as such, is oblivious to novel (or simply alternative) aspects of the situation.”³² The problem with this approach is that it implies a rigid distinction between mindfulness and mindlessness, whereby only the latter leads to deception. When users interact with AI, however, they also replicate social behaviors and habits in self-conscious and reflective ways. Users, for instance, carry out playful exchanges with AI voice assistants, asking Siri or Alexa if they are human. They wish them goodnight before going to bed, even if aware that the AI assistants will not ‘sleep’ in the same sense as humans do.³³ This suggests that distinctions between mindful and mindless behaviors fail to capture the complexity of the interaction. In contrast, obliviousness implies that while deception is not thematized as such by the user, she or he may engage in social interactions with the machine deliberately as well as unconsciously. Obliviousness also allows the user to maintain at least the illusion of control – this being, in the age of user friendliness, a key principle of software design.³⁴

The fourth characteristic of banal deception is its low definition. While this term is commonly used to describe formats of video or sound reproduction with lower resolution, in media theory the term has also been employed in reference to media that demand more participation from audiences and users in the construction of sense and meaning.³⁵ For what concerns AI, textual and voice interfaces are low definition because they leave ample space for

the user to imagine and attribute characteristics such as gender, race, class and personality to the disembodied voice or text. For instance, voice assistants do not present at a physical or visual level the appearance of the virtual character (such as “Alexa” or “Siri”), but some cues are embedded in the sounds of their voices, in their names and in the content of their exchanges. It is for this reason that, as shown in research about people’s perceptions of AI voice assistants, different users imagine AI assistants in different, multiple ways, which also enhances the effect of technology being personalized to each individual.³⁶ In contrast, humanoid robots leave less space for the users’ imagination and projection mechanisms, and are therefore not low definition. This is why disembodied AI voice assistants have become much more influential today than humanoid robots: the fact that users can project their own imaginations and meanings makes interactions with these tools much more personal and reassuring, and therefore it is easier to incorporate them than robots into our everyday lives.³⁷

The fifth and final defining characteristic of banal deception is that it is not just imposed on users, but also ‘programmed’ by designers and developers. This is why the word “deception” is preferable to illusion, since deception implies some form of agency, permitting clearer acknowledgement of how developers of AI technologies work towards achieving the desired effects. In order to explore and develop the mechanisms of banal deception, designers need to project an image of the expected user. In actor-network theory this corresponds to the notion of script, which refers to the work of innovators as ‘inscribing’ visions or predictions about the world and the user in the technical content of the new object and technology.³⁸ Although this is always an exercise of imagination, it draws on specific efforts to gain knowledge about users, or more generally about ‘humans’. Recent work in human-computer interaction acknowledges that “perhaps the most difficult aspect of interacting with humans is the need to model the beliefs,

desires, intentions preferences, and expectations of the human and situate the interaction in the context of that model.”³⁹ The historical excavation undertaken in this book shows that this work of modelling users is as old as AI itself. As soon as interactive systems were developed, computer scientists and AI researchers explored how human perception and psychology functioned and attempted to use such knowledge to close the gap between computer and users.⁴⁰

It is important to stress that considering the agency of programmers and developers who design and prepare for use AI systems is perfectly compatible with the recognition that users have themselves agency. As much critical scholarship on digital media shows, in fact, users of digital technologies and systems often subvert and reframe the intentions and expectations of companies and developers.⁴¹ This does not imply, however, that the latter do not have an expected outcome in mind. As Taina Bucher recently remarked, “the cultural beliefs and values held by programmers, designers, and creators of software matter”: we should examine and question their intentions despite the many difficulties involved in reconstructing them retrospectively from the technology and its operations.⁴²

Importantly, the fact that banal deception is not to be seen as negative by default does not mean that its dynamics should not be the subject of attentive critical inquiry. One of the key goals of this book is to identify and counteract potentially problematic practices and implications that emerge as a consequence of the incorporation of banal deception into AI. Unveiling the mechanisms of banal deception, in this sense, is also an invitation to interrogate what the “human” means in the discursive debates and the practical work that shape the development of AI. As the trajectory described in this book demonstrates, the modelling of the “human” developed throughout the history of AI has in fact been quite limited. Even when computer access was progressively extended to wider potential publics, the expected user was often

envisioned as a white, educated man, perpetuating biases that remain inherent in contemporary computer systems.⁴³ Furthermore, studies and assumptions about how users perceive and react to specific representations of gender, race and class were implemented in interface design, leading for instance to gendered characterizations of many contemporary AI voice assistants.⁴⁴

One further issue is the extent to which the mechanisms of banal deception embedded in AI are changing the social conventions and habits that regulate our relationships with both humans and machines. Pierre Bourdieu uses the concept of *habitus* to characterize the range of dispositions through which individuals perceive and react to the social world.⁴⁵ Since *habitus* is based on previous experiences, the availability of increasingly more opportunities to engage in interactions with computers and AI is likely to feed forward into our social behaviors in the future. The subtitle of this book refers to social life *after* the Turing Test, but even if a computer program able to pass the Test is still to be created, the dynamics of banal deception in AI already represent an inescapable influence on the social life of millions people around the world. The main objective of this book is to neutralize the obliviousness of banal deception, bringing its mechanisms to the surface to better understand new AI systems that are altering societies and everyday life.

AI, communication, media history

AI is a highly interdisciplinary field, characterized by a range of different approaches, theories and methods. Some AI-based applications, such as the information-processing algorithms that regulate access to the Web, are a constant presence in the everyday lives of masses of people; others, like industrial applications of AI in factories and workshops, are rarely, if ever, encountered.⁴⁶ This book focuses particularly on communicative AI, i.e. AI applications that are

designed to enter into communication with human users.⁴⁷ Communicative AIs include applications involving conversation and speech, such as natural language processing (NLP), chatbots, social media bots and AI voice assistants. The field of robotics makes use of some of the same technologies developed for communicative AI – for instance to have robots communicate through a speech dialogue system – but remains outside the remit of this book. As Andreas Hepp has recently argued, in fact, AI is less common today as embodied physical artefacts than they are as software applications.⁴⁸ This circumstance, as mentioned above, may be explained by the fact that computers do not match one of the key characteristics of banal deception: low definition.

Communicative AI departs from the historical role of media as mere channels of communication, since AI also acts as producer of communication, with which humans (as well as other machines) exchange messages.⁴⁹ Yet communicative AI is still a medium of communication, and therefore inherits many of the dynamics and structures that have characterized mediated communication at least since the emergence of electronic media in the nineteenth century. This is why, to understand new technologies such as AI voice assistants or chatbots, it is vital to contextualize them within the history of media.

As communication technologies, media draw from human psychology and perception, and it is possible to look at media history in terms of how deceitful effects were incorporated within different media technologies. Cinema achieves its effects by exploiting the limits of human perception, such as the impression of movement that can be given through the fast succession of a series of still images.⁵⁰ Similarly, as Jonathan Sterne has aptly shown, the development of sound media drew from knowledge about the physical and psychological characteristics of human hearing and listening.⁵¹ In this sense, the key event of media history

since the nineteenth century was not the invention of any new technology such as the telegraph, photography, cinema, television or the computer. It was instead the emergence of the new human sciences, from physiology and psychology to the social sciences, which provided the knowledge and epistemological framework to adapt modern media to the characteristics of the human sensorium and intellect.

Yet, the study of media has often fallen in the same trap of those who believe that deception in AI matters only if it is “deliberate” and “straight-up.”⁵² Deception in media history has mainly been examined as an exceptional circumstance, highlighting the manipulative power of media rather than acknowledging deception’s structural role in modern media. According to an apocryphal but persistent anecdote, for instance, early movie audiences exchanged representation for reality and panicked before the image of an incoming train.⁵³ Similarly, in the story of Orson Welles’ “War of the Worlds” radio broadcast that reportedly many interpreted as a report of an actual extraterrestrial invasion, live broadcasting has led people to confuse fiction with reality.⁵⁴ While such blatant (and often exaggerated) cases of deception have attracted much attention, few have reflected on the fact that deception is a key feature of media technologies function – that deception, in other words, is not an incidental but an irremediable characteristic of media technologies.⁵⁵

To uncover the antecedents of AI and robotics, historians commonly point to automata, self-operating machines mimicking the behavior and movements of humans and animals.⁵⁶ Notable examples in this lineage include the mechanical duck built by French inventor Jacques de Vaucanson in 1739, which displayed the ability of eating, digesting and defecating, and the Mechanical Turk that amazed audiences in Europe and America in the late eighteenth and early nineteenth century with its proficiency at playing chess.⁵⁷ In considering the relationship

between AI and deception, these automata are certainly a case in point, as their apparent intelligence was the result of manipulation by their creators: the mechanical duck had feces stored in its interior, so that no actual digestion took place, while the Turk was maneuvered by a human player hidden inside the machine.⁵⁸ I argue, however, that to fully understand the broader relationship between contemporary AI and deception, one needs to delve into a wider historical context that goes beyond the history of automata and programmable machines. This context is the history of deceitful media, i.e. of how different media and practices, from painting and theatre to sound recording, television and cinema, have integrated banal deception as a strategy to achieve particular effects in audiences and users. Following this trajectory shows that some of the dynamics of communicative AI are in a relationship of continuity with the ways in which audiences and users projected meaning onto other media and technology.

Examining the history of communicative AI from the proposal of the Turing Test in 1950 to the present day, I move from the persuasion that a historical approach to media and technological change helps us comprehend ongoing transformations in the social, cultural and political spheres. Scholars such as Lisa Gitelman, Erkki Huhtamo and Jussi Parikka have compellingly shown that what we call “new media” have a long history, whose study is necessary to understand today’s digital culture.⁵⁹ If it is true that history is one of the best tools at our disposal to comprehend the present, I believe that it is also one of our best instruments, although still an imperfect one, to anticipate the future. In areas of rapid development such as AI, it is extremely difficult to forecast even short- and medium-term development, let alone long-term changes.⁶⁰ Looking at longer historical trajectories across several decades helps us identify key trends and trajectories of change that have characterized the field across several decades and might, therefore, continue to shape it in the future. Although it is important to understand how

recent innovations like neural networks and deep learning work, we also need a better sense of the directions through which the field has moved across a longer timeframe. Media history, in this sense, is a science of the future: it does not only shed light on the dynamics by which we arrived where we are today, but helps pose new questions and problems through which we may navigate the technical and social challenges ahead.⁶¹

Following Lucy Suchman, I use the terms ‘interaction’ and ‘communication’ interchangeably, since interaction entails the establishment of communication between different entities.⁶² Early approaches in human-computer interaction (HCI) recognized that interaction was always intended as a communicative relationship, and the idea that the computer is both a channel and a producer of communication is much older than often implied.⁶³ Although AI and HCI are usually framed as separate, considering them as distinct limits our capacity to understand their development and impact. Since the very origins of their field, AI researchers have reflected on how computational devices could enter into contact and dialogue with human users, bringing the problems and questions relevant to HCI to the center of their own investigation. Exploring the intersections between these fields helps understand that they are united by a key tenet: the fact that when a user interacts with technology, the responsibility for the outcome of such interaction is shared between the technology and the human.

On a theoretical level, the book is indebted to insights from different disciplinary fields, from action-network theory to social anthropology, from media theory to film studies and art history. I use these diverse frameworks as tools to propose an approach to AI and digital technologies that emphasizes humans’ participation in the construction of meaning. As works in actor-network theory as well as social anthropologists such as Armin Appadurai and Alfred Gell have taught us, not only humans, but also artifacts can be regarded as social agents in particular

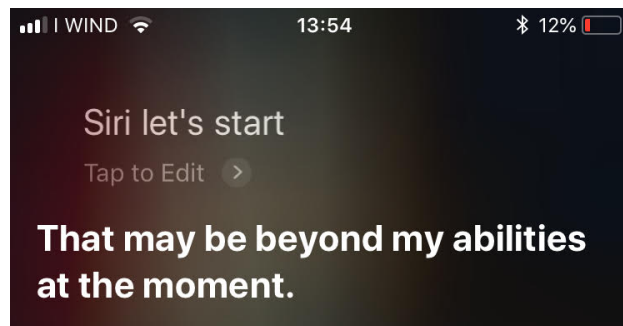
social situations.⁶⁴ People often attribute intentions to objects and machines: for instance, car owners attribute personalities to their cars and children to their dolls. Things, like people, have social lives, and their meaning is continually negotiated and embedded within social relations.⁶⁵

In media studies, the implications of this discovery have been examined by scholars that have shifted from decade-long reflections on the audiences of media such as radio, cinema and television to developing a new focus on the interactive relationships between computers and users. In *The Media Equation*, a foundational work published in the mid-1990s, Byron Reeves and Clifford Nass argue that we tend to treat media, including but not only computers, in accordance with the rules of social interaction.⁶⁶ Later studies by Nass, Reeves and other collaborators established what is known as the Computer Are Social Actors (CASA) paradigm, which contends that humans apply social rules and expectations to computers, and have explored the implications of new interfaces that talk and listen to users, which are becoming increasingly available in computers, cars, call centres, domestic environments and toys.⁶⁷ Another crucial contribution to such endeavors is that of Sherry Turkle. Across several decades, her research has explored interactions between humans and AI, emphasizing how their relationship does not follow from the fact that computational objects really have emotions or intelligence, but from what they evoke in their users.⁶⁸

Although the role of deception is rarely acknowledged in discussions of AI, I argue that interrogating the ethical and cultural implications of such dynamics is an urgent task that needs to be approached through interdisciplinary reflection at the crossroads between computer science, cognitive science, social sciences and the humanities. While the public debate on the future of AI tends to focus on the hypothesis that AI will make computers as intelligent or even more intelligent than people, we also need to consider the cultural and social consequences of

deceitful media providing the appearance of intelligence. In this regard, the contemporary obsession with apocalyptic and futuristic visions of AI, such as singularity, superintelligence and the robot apocalypse, makes us less aware of the fact that the most significant implications of AI systems are to be seen not in a distant future, but in our ongoing interactions with ‘intelligent’ machines.

Technology is shaped by the agency of scientists, designers, entrepreneurs, users and policy-makers, but also by the kinds of questions we ask about it. *Deceitful Media* hopes to inspire readers to ask new questions about the relationship between humans and machines in today’s world. We will have to start searching for answers ourselves, as the “intelligent” machines we are creating can offer no guidance on such matters. As one of those machines admitted:



Endnotes

¹ ‘Google’s AI Assistant Can Now Make Real Phone Calls’, 2018, available at

https://www.youtube.com/watch?v=JvbHu_bVa_g&time_continue=1&app=desktop (retrieved 12 January 2020). See also O’Leary, ‘Google’s Duplex’.

² As critical media scholar Zeynep Tufekci put it in a tweet that circulated widely. The full thread on Twitter is readable at <https://twitter.com/zeynep/status/994233568359575552> (retrieved 16 January 2020).

³ Joel Hruska, ‘Did Google Fake Its Duplex AI Demo?’, *ExtremeTech*, 18 March 2018, available at <https://www.extremetech.com/computing/269497-did-google-fake-its-google-duplex-ai-demo> (retrieved 16 January 2020).

⁴ See, among many others, on the one side, Minsky, ‘Artificial Intelligence’; Kurzweil, *The Singularity Is Near*; and on the other side, Dreyfus, *What Computers Can’t Do*; Smith, *The AI Delusion*.

⁵ Smith and Marx, *Does Technology Drive History?*; Williams, *Television*; Jones, ‘The Technology Is Not the Cultural Form?’

⁶ Goodfellow, Bengio and Courville, *Deep Learning*.

⁷ Benghozi and Chevalier, ‘The Present Vision of AI... or the HAL Syndrome’.

⁸ See Chapter 2.

⁹ Gombrich, *Art and Illusion*.

¹⁰ There are, however, some relatively isolated but still significant exceptions. Adar et al., for instance, distinguish between malicious and benevolent deception, which they describe as “deception aimed at benefitting the user as well as the developer.” Such a benevolent form of deception, they note, “is ubiquitous in real-world system designs, although it is rarely described

in such terms.” Adar, Tan and Teevan, ‘Benevolent Deception in Human Computer Interaction’. Similarly, Chakraborti and Kambhampati observe that the obvious outcome of embedding models of mental states of human users into AI programs is that it opens up the possibility of manipulation. Chakraborti and Kambhampati, ‘Algorithms for the Greater Good!’. From a different perspective, Naeve and Grabowski have conceptualized communications between human and machine as “pseudop-communicaton,” arguing for the importance of a semiotic perspective to understand human-computer interaction. Naeve and Grabowski, ‘Human–computer interaction viewed as pseudo-communication’. See also Castelfranchi and Tan, *Trust and Deception in Virtual Societies*.

¹¹ Coeckelbergh, ‘How to Describe and Evaluate “Deception” Phenomena’; Schuetzler, Grimes and Giboney, ‘The Effect of Conversational Agent Skill on User Behavior during Deception’; Tognazzini, ‘Principles, Techniques, and Ethics of Stage Magic and Their Application to Human Interface Design’.

¹² DePaulo et al., ‘Lying in Everyday Life’; Steinel and De Dreu, ‘Social Motives and Strategic Misrepresentation in Social Decision Making’; Solomon, ‘Self, Deception, and Self-Deception in Philosophy’; Barnes, *Seeing through Self-Deception*. A broad, traditional definition of deception is “to cause to believe what is false”; see Mahon, ‘The Definition of Lying and Deception’.

¹³ Acland, *Swift Viewing*.

¹⁴ Martin, *The Philosophy of Deception*, 3; Rutschmann and Wiegmann, ‘No Need for an Intention to Deceive?’.

¹⁵ Wrathall, *Heidegger and Unconcealment*, 60.

¹⁶ Hoffman, *The Case Against Reality*.

¹⁷ Pettit, *The Science of Deception*.

¹⁸ Hyman, *The Psychology of Deception*. Notable examples of early studies on deception from a psychology perspective include Triplett, 'The Psychology of Conjuring Deceptions'; Jastrow, *Fact and Fable in Psychology*.

¹⁹ Parisi, *Archaeologies of Touch*; Littlefield, *The Lying Brain*; Alovio, 'Lo schermo di Zeusi'; Sterne, *The Audible Past*.

²⁰ Martin, *The Philosophy of Deception*, 3

²¹ Caudwell and Lacey, 'What Do Home Robots Want?'

²² The attribute of "banal" has been applied in the social sciences in regards with other topics, more famously by Michael Billig with his insightful theorization of "banal nationalism."

Although there are some points of analogies with the concept of banal deception, especially regarding the hidden character and the fact that the banal is often underscored and not interrogated as such, overall banal nationalism has been a useful source of inspiration rather than a direct reference to develop my theory. See Billig, *Banal Nationalism*; as well as Hjarvard, 'The Mediatization of Religion', which proposes the notion of "banal religion" in specific reference to Billig's work.

²³ See, among others, Guzman, 'Imagining the Voice in the Machine'; Reeves and Nass, *The Media Equation*; Turkle, *The Second Self*.

²⁴ Guzman, 'Beyond Extraordinary'.

²⁵ Ekbja, *Artificial Dreams*; Finn, *What Algorithms Want*.

²⁶ Guzman, 'Beyond Extraordinary', 84.

²⁷ Papacharissi, *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*.

²⁸ Chun, 'On "Sourcery," or Code as Fetish'.

²⁹ Black, 'Usable and Useful'.

³⁰ Porcheron Fischer, Reeves and Sharples ‘Voice Interfaces in Everyday Life’; Guzman, *Imagining the Voice in the Machine*.

³¹ Nass and Moon, ‘Machines and Mindlessness’; Kim and Sundar, ‘Anthropomorphism of Computers’.

³² Langer, ‘Matters of Mind’.

³³ Guzman, ‘Making AI Safe for Humans’.

³⁴ Black, ‘Usable and Useful’.

³⁵ Ortoleva, *Miti a bassa intensità*. Marshall McLuhan employs the notion of “cool media” to describe such media that require more participation from users or audiences. McLuhan, *Understanding Media*.

³⁶ Guzman, *Imagining the Voice in the Machine*.

³⁷ Hepp, ‘Artificial Companions, Social Bots and Work Bots’. On discussions of perceptions of robots, see the concept of the “uncanny valley,” originally proposed in Mori, ‘The Uncanny Valley’ (the original text in Japanese was published in 1970). It is also worth noting that AI voice assistants are “disembodied” only to the extent that they are not given a proper physical ‘body’ whose movements they control, as in robots; however, all software has to some extent its own materiality, and AI voice assistants in particular are always embedded in material artefacts such as smartphones or smart speakers. See, on the materiality of software, Kirschenbaum, *Mechanisms*; and specifically on AI voice assistants, Guzman, ‘Voices in and of the Machine’.

³⁸ Akrich, ‘The De-Description of Technical Objects’. See also Feenberg, *Transforming Technology*; Forsythe, *Studying Those Who Study Us*.

³⁹ Chakraborti and Kambhampati, ‘Algorithms for the Greater Good!’.

⁴⁰ Interestingly, deception plays a key role not only in how knowledge about human psychology and perception is used, but also in how it is collected and accumulated. Forms of deception, in fact, have played a key role in experimental design of psychological studies to mislead participants so that they remained unaware of the actual purposes of an experiment. See Korn, *Illusions of Reality*.

⁴¹ Balbi and Magaudda, *A History of Digital Media*.

⁴² Bucher, *If... Then*, 68.

⁴³ Towns, 'Towards a Black Media Philosophy'.

⁴⁴ Sweeney, 'Digital Assistants'.

⁴⁵ Bourdieu, *Outline of a Theory of Practice*.

⁴⁶ Guzman, 'The Messages of Mute Machines'.

⁴⁷ For a clear and compelling description of the remit of "communicative AI", see Guzman and Lewis, 'Artificial Intelligence and Communication'.

⁴⁸ Hepp, 'Artificial Companions, Social Bots and Work Bots'.

⁴⁹ Guzman, *Human-Machine Communication*; Gunkel, 'Communication and Artificial Intelligence'; Guzman and Lewis, 'Artificial Intelligence and Communication'. For a discussion of the concept of medium in AI and communication studies, see Natale, 'Communicating with and Communicating through'.

⁵⁰ Doane, *The Emergence of Cinematic Time*; Hugo Münsterberg, *The Film: A Psychological Study*.

⁵¹ Sterne, *The Audible Past*; Sterne, *MP3*. Even literature has been examined by prominent writers, such as Edgar Allan Poe in his "Philosophy of Composition," for its capacity to achieve

psychological effects through specific stylistic means. Poe, *The Raven*; with, *The Philosophy of Composition*.

⁵² As shown, for instance, by the controversy about Google Duplex mentions at the beginning of this introduction.

⁵³ Bottomore, ‘The Panicking Audience?’; Martin Loiperdinger, ‘Lumière’s Arrival of the Train’; Sirois-Trahan, ‘Mythes et Limites Du Train-Qui-Fonce-Sur-Les-Spectateurs’.

⁵⁴ Pooley and Socolow, ‘War of the Words’; Heyer, ‘America under Attack I’; Hayes and Battles, ‘Exchange and Interconnection in US Network Radio.

⁵⁵ There are however important exceptions; among works of media history that make important contributions to the study of deception and media, it is worth mentioning Sconce, *The Technical Delusion*; Acland, *Swift Viewing*.

⁵⁶ See, for instance, McCorduck, *Machines Who Think*; Boden, *Mind as Machine*

⁵⁷ Riskin, ‘The Defecating Duck, or, the Ambiguous Origins of Artificial Life’; Sussman, ‘Performing the Intelligent Machine’; Cook, *The Arts of Deception*.

⁵⁸ Geoghegan, ‘Visionäre Informatik’. Sussman, ‘Performing the Intelligent Machine’.

⁵⁹ Gitelman, *Always Already New*; Huhtamo, *Illusions in Motion*; Parikka, *What Is Media Archaeology?*

⁶⁰ This is why predictions about technology so often fail. See on this Ithiel De Sola Pool et al., ‘Foresight and Hindsight’; Natale, ‘Introduction: New Media and the Imagination of the Future’.

⁶¹ Park, Jankowski and Jones, *The Long History of New Media*; Balbi and Magaudda, *A History of Digital Media*.

⁶² Suchman, *Human-Machine Reconfigurations*.

⁶³ For an influential example, see Licklider and Taylor, 'The Computer as a Communication Device'.

⁶⁴ Appadurai, *The Social Life of Things*; Gell, *Art and Agency*; Latour, *The Pasteurization of France*.

⁶⁵ Edwards, 'Material Beings'.

⁶⁶ Reeves and Nass, *The Media Equation*.

⁶⁷ See, among many others, Nass and Brave, *Wired for Speech*; Nass and Moon, 'Machines and Mindlessness'.

⁶⁸ Turkle's most representative works are *Reclaiming Conversation*; *Alone Together*; *Evocative Objects*; *The Second Self*; *Life on the Screen*.