

Subscribe

Sign  
Iniai  
Playeriai  
NewsIAI  
LiveIAI  
Academyiai  
Podcast

More ▾

Search...



# Changing How the World Thinks

An online magazine of big ideas

Home

Philosophy

Science

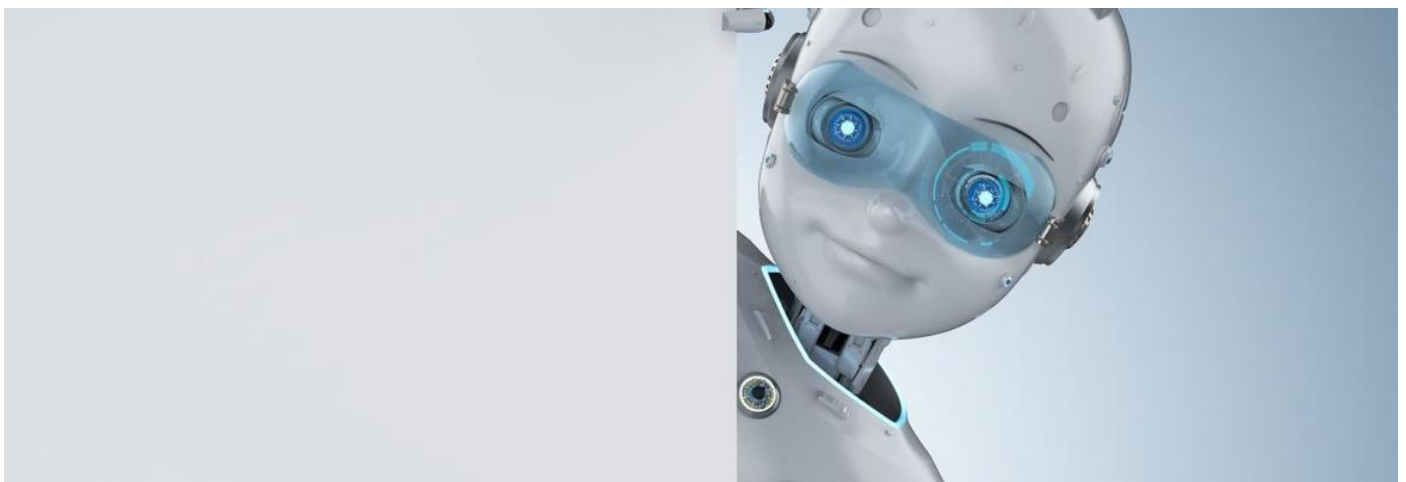
Politics

Arts

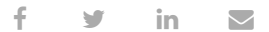
Issue Archive

# Artificial intelligence and gullible humans

The Turing Test and the real significance of AI



Issue 96, 10th May 2021



**Simone Natale** | Author of *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*

**1,847 words**

Read time: approx. 9 mins

*The real significance of new AI technology is not that machines can be like humans but that humans are prone to deception, argues Simone Natale.*

In the last few weeks, the official [Twitter account](#) of the Perseverance Mars Rover, a car-sized robot designed by NASA to explore Mars, attracted news media attention and a steadily growing number of followers. While many marvelled at the photos of the Red Planet that the account posted, some asked if the fact that the account is tweeting in the first person - as if it was the Rover itself reporting from Mars and not NASA's PR office – [amounts to a form of deception](#). Is anthropomorphization an honest way to communicate the rover's functioning to the public? Are we being led to project consciousness and sociality onto a machine that has neither?



Similar questions about Artificial Intelligence (AI) are being posed more



### The uncontrollability of Artificial Intelligence

[Read more](#)

and more frequently. As AI technologies become more pervasive and influential, many fear that it will become more difficult to distinguish between “real” AI technologies and blatant frauds. They point to AI and robotics companies using marketing tools and design features that exaggerate the apparent intelligence of robots. This debate, however, misses an important point. If we want to really understand the social and cultural

dynamics activated by the new generation of AI and robots, we need to acknowledge that deception is not an incidental feature of these technologies. It is not, in other words, something that only characterizes certain uses and expressions of AI technology. Deception is, instead, ingrained in the very essence of what AI is and how it works. It is as central to AI as the circuits and software that make it run.

### AI and deception

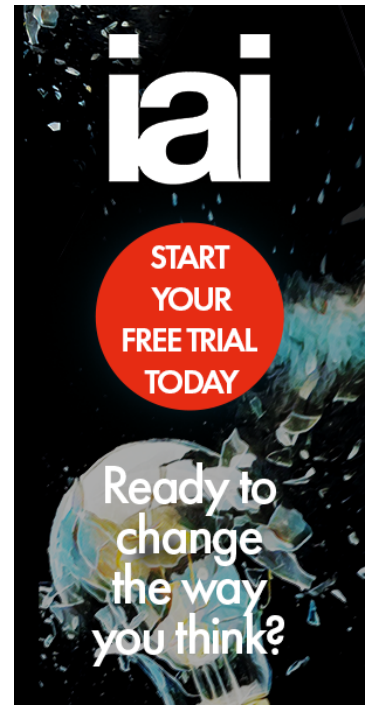
Because the term “deception” is usually associated with malicious endeavours, the AI and computer science communities are usually hesitant to discuss their work in terms of deception, unless it is presented as an unwanted, exceptional outcome that only characterises specific objects and situations. Such an approach, however, relies on a rigid understanding of deception that does not sit well with recent explorations of this concept.

## AI scientists collected information on how users react to machines exhibiting the appearance of intelligent behaviours and incorporated this knowledge into the design of software and machines.

Scholars in social psychology, philosophy, and sociology have shown that deception is an inescapable fact of social life and plays a key role in social interaction and communication. [As philosopher Mark Wrathall put it](#), “it rarely makes sense to say that I perceived either truly or falsely,” since deception is an integral part of how we perceive and navigate the world. 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 – our tendency and ability to identify patterns in visual information – would have helped me, on another occasion, to identify a potential danger. This shows us how deception is functional to our ability to navigate the external world.

[In my own work](#), I demonstrate that computer scientists have worked since the early history of their field to exploit the limits and affordances of our perception and intellect. AI scientists collected information on how users react to machines exhibiting the appearance of intelligent behaviours and incorporated this knowledge into the design of software and machines. For instance, robotics engineers soon realised that variations in the external appearance of robots could invite specific feelings in users. A lively debate ensued about the ways users would react to different kinds of design. The results of this discussion are visible in home robots such as [Jibo](#), whose design is meant to awaken sentiments of empathy in their owners.

### The Turing Test, or how a machine can trick us



Get the **big ideas** straight to your inbox weekly

Enter Email Address

Sign me up

### More from this issue:

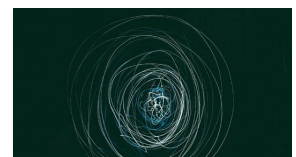


Infinite possibilities, infinite worlds



The self in the cloud

### Related Videos:



The close relationship between deception and AI was recognised by one of the earliest and most perceptive pioneers of this field, the British mathematician Alan Turing, who proposed the Imitation Game, now more commonly called “Turing Test,” in a paper published in 1950. Turing started his paper by asking if machines can think, only to dismiss the question as useless, since it would be impossible, he reasoned, to find agreement on what “thinking” means. He proposed, therefore, to substitute that question with a practical experiment, the Turing Test, in which a human judge enters into conversation with an interlocutor through written messages, as one can do today in a chatroom. The judge has to find out if her or his conversation partner is a human or a machine. A computer program will then pass the Test if it proves able to convince the judge that it is a human.

Usually, the Turing Test is discussed as a threshold to assess AI. What is really interesting about the Test, however, is not whether computers will be able to pass it, but rather how the Test overturns our viewpoint on AI. The implication of the Turing test is that we shouldn’t seek an absolute definition of AI, but instead, we should define AI from the point of view of the observer.

Deception plays a central role in this context: the machine passes the test if it is able to deceive a human judge. Put it in other words, AI is AI if it convinces us of being so.

**Physicist Roger Penrose, computer scientist Nigel Shadbolt and novelist and digital age icon Warren Ellis consider the threat of intelligent machines.**

### The banality of deception

Some might object that the situation described in the Turing Test, where a machine tries to deceive us into believing it is human, doesn’t reflect our everyday experience with AI. It might happen in some specific, relatively rare situations – for instance with some bots on social media – but is not key to our most common uses and engagements with AI.

This, however, doesn’t mean that more subtle but still significant forms of deception are not at play in everyday interactions with AI technologies. Take, for instance, the case of voice assistants such as Apple’s Siri or Amazon’s Alexa. The choice to assign to voice assistants humanlike voices instead of synthetic-sounding voices, and to provide them with precise connotations of gender and even social and regional accent, are never made by chance. They derive from considerations made by Amazon and Apple on how users will react to different kinds of voices, and how such reactions can help achieve specific outcomes – for instance, encouraging users to integrate these tools into their domestic environments and everyday lives. Although users of Siri and Alexa are perfectly able to understand that these are computer programs and not real persons, the gendered voice

## Can We Control AI?



## Zanna Clay | In-depth Interview



## How to talk about climate change



## How to live forever



## The smartphone society

of virtual assistants activates mechanisms of representation by which users imagine a source for the voice – and, subsequently, a stable character with which to interact, even if within relatively strictly boundaries. The gender, class and race clues embedded in the synthetic voices create the psychological and social conditions for projecting an identity and, to some extent, a personality onto the virtual assistant.

We can describe these forms of apparently inoffensive deception with the term “[banal deception](#).” Banal deception entails mundane, everyday situations in which technologies and devices mobilize specific elements of user’s perception and psychology – such as the all-too-human tendency to attribute agency to things or personality to voices. The deception of Alexa and Siri is “banal” because it has to do with situations that are immersed in our everyday life and that we don’t even perceive as deceptive. The ordinary and mundane character of banal deception makes it unnoticeable but more consequential since it helps these technologies enter the deepest layers of our everyday habits and behaviours.

In contrast to outright, ‘stronger’ forms of deception, banal deception can have, at least potentially, some value for the user. For example, the fact that users respond socially to voice assistants brings an array of pragmatic benefits: it makes it easier to use these tools and creates space for playful interaction and emotional reward. For companies and designers, moreover, the fact that banal deception is not perceived as such has an advantage at a commercial level, since the user maintains the illusion of having full control of the experience.

Examples of banal deception in contemporary AI systems are manifold. Voice assistants [mobilize gender and class stereotypes](#) through the accent and the characterization of their voice, aiming at specific responses from users. Similar mechanisms are activated by chatbots in written communications, as [shown by studies](#) of users’ responses to chatbots using emojis. In companion robots and chatbots, [a sense of cuteness](#) is created through specific aspects of their design, helping to activate mechanisms of empathy that provide an emotional reward to users. [On social media](#), bots programmed to impersonate fictional characters populate accounts that disclose their mechanical nature, but still stimulate engagement from users.

**The ordinary and mundane character of banal deception makes it unnoticeable but more consequential since it helps these technologies enter the deepest layers of our everyday habits and behaviours.**

### The machine and the human

The notion of banal deception helps identify that deception is not an exception in AI; on the contrary, it plays a fundamental role in AI technologies that are programmed to interact with humans. One of the implications of this is that AI system cannot be understood only by examining the internal functioning of the machine: one needs to inquire also the social, psychological, and cultural dynamics that AI activates when it engages users.

Public discussions of AI usually emphasise the evolution of the technology, which has become more and more sophisticated and performative. But in fact, the functioning of AI and its impact does not depend only on technical features. As Alan Turing intuited when he proposed his Test, the outcome of the interaction between a machine and a human depends on the characteristics of the machine as well as on the characteristics of the humans who participate in the interaction. For instance, if a judge of the Turing Test has a strong background in computer science, she or he will likely give a different assessment than someone who hasn’t much knowledge or experience with this technology. The result is that developments in AI have to do with technical advances as well as with the development of a range of features and strategies by which AI developers mobilise

perceptive mechanisms, habits, and social understandings of human users in order to make these technologies achieve the desired effect.

The questions posed in the debate about Perseverance’s Twitter account should thus be reformulated. What we should ask, instead, is how to develop more informed and self-aware relationships with technologies that are programmed to take advantage of our liability to be deceived. It might sound paradoxical, but to better comprehend AI we need first to better comprehend ourselves. Contemporary AI technologies constantly mobilize mechanisms such as empathy, stereotyping, and social habits. To understand these technologies more deeply, and to fully appreciate the relationship we are building with them, we need to interrogate how such mechanisms work and which part deception plays in our interaction with “intelligent” machines.

**Simone Natale**  
Issue 96, 10th May 2021



LATEST RELEASES



**Getting real about global power**



**The last mystery of cosmology**



**You don't need an identity**

**Join the conversation**

**Sign in** to post comments or **join now (only takes a moment)**. Don't have an account? Sign in with Facebook, Twitter or Google to get started:



© The Institute of Art and Ideas