From human intelligence to artificial smartness

The undone science of artificial intelligence

Alexandre Kabbach*

University of Geneva

In defining *artificial intelligence*, Wang (2019) stresses that the fundamental design principle that separates machines from human minds is that "a program is traditionally designed to do something in a predetermined *correct* way, while the mind is constructed to *do its best* using whatever it has" (Wang, 2019, p.16). In the context of the Turing Test (Turing, 1950) and the original search for human intelligence, this "correctness criterion" is usually considered to apply to *human behavior*, so that machines could be said to be designed so as to *follow a norm*. This concept of "norm" here is to be understood as a "norm of humanness", where to behave "correctly" for a machine means to behave "humanly"; to satisfy a set of constraints that would make its behavior indistinguishable from that of another human. My question, then, is the following: does that assumption still holds today with modern artificial intelligence systems such as ChatGPT (Ouyang et al., 2022)? That is, is it enough for such models to equate *correct* behavior with *human* behavior? I will argue no.

In this conference, I propose to argue that the practical applications under which modern artificial intelligence systems are put to use have diverted us from the original goal of modeling human intelligence. As far as modern artificial intelligence is concerned indeed, human *intelligence* is just not enough: machines need to be *smart*.

My contribution relies on a fundamental distinction between (human) *intelligence* and *smartness*. Following the original conception of (Turing, 1950), I argue that "intelligence" is a human *faculty*; an ability to "think" that one possesses by virtue of being human (see Shieber, 2004, footnote 2, p.6). "Smartness", on the other hand, is a *normative ideal*; a specification of how human beings *ought* to behave rather than how they *do* behave in practice. My argument, then, is that human intelligence is fundamentally different from smartness, for *human beings "make mistakes"*: they always deviate from whichever normative ideal of smartness they live by one way or another. To illustrate my argument, I propose to consider the example of *spelling* which I treat as a prototypical case of a normative ideal on (linguistic) behavior that members of a particular linguistic community ought to abide to but never quite manage to in practice. Even the best speller, I would argue, is bound to violate the norms of spelling and "make mistakes" at some point. I put that argument into perspective with Turing's original answer to the "Arguments from Various Disabilities"—telling us precisely that machines would have to make deliberate mistakes so as to appear humans (Turing, 1950, pp.447–449)—and turn to reports of practical Turing Tests in (British) English which have shown specifically that human judges rely on spelling mistakes made by human participants to distinguish them from machines (see Warwick & Shah, 2016, p.1001).

I conclude that we must come to term with the idea that researching *intelligence* and *smartness* are just two different and potentially irreconcilable scientific endeavors, and that there are no reasons for us to expect "smart machines" such as ChatGPT to get us any closer to a proper understanding of human intelligence. After decades

*alexandre@kabbach.net

of experiencing machines that were just "too bad" to be human, my intuition indeed is that we are progressively shifting to a situation where machines will just prove "too good" to be human this time, but in any case equally unable to pass the Turing Test and/or to display any form of human intelligence. Moreover, at a time where "government funding is being eclipsed by consumer markets" (Church, 2018, p.1) and where the priorities of the field of artificial intelligence research are becoming more and more dictated by industrial needs (Ahmed et al., 2023), we must ask ourselves whether there is really any money to be made by those industries—or power to be gained—by building systems that "make mistakes" and deviate from whichever normative ideal of smartness they are being put to use. Why would one deliberately create a language model that makes spelling mistakes indeed, especially if they can actually avoid it?

I then propose to open up the discussion by returning to Turing's original answer to the "Arguments from Various Disabilities" in order to argue that a proper understanding of human intelligence requires a proper understanding of *human subjectivity*, that is, of what distinguishes a *human* from a *non-human* deviation from the normative ideal of smartness. In other words, I argue that, to properly understand what human intelligence is actually made of, we need to understand precisely *why* we "make mistakes" in the first place, and what distinguishes a *human* "mistake" from a *non-human* "mistake".

References

Ahmed, Nur & Wahed, Muntasir & Thompson, Neil C. 2023. The growing influence of industry in AI research. *Science* 379(6635). 884–886. doi:10.1126/science.ade2420. https://doi.org/10.1126/science.ade2420

Church, Kenneth Ward. 2018. Emerging trends: A tribute to Charles Wayne. *Natural Language Engineering* 24(1). 155–160. doi:10.1017/S1351324917000389. https://doi.org/10.1017/S1351324917000389

Ouyang, Long & Wu, Jeffrey & Jiang, Xu & Almeida, Diogo & Wainwright, Carroll & Mishkin, Pamela & Zhang, Chong & Agarwal, Sandhini & Slama, Katarina & Ray, Alex & Schulman, John & Hilton, Jacob & Kelton, Fraser & Miller, Luke & Simens, Maddie & Askell, Amanda & Welinder, Peter & Christiano, Paul F. & Leike, Jan & Lowe, Ryan. 2022. Training language models to follow instructions with human feedback. In Koyejo, S. & Mohamed, S. & Agarwal, A. & Belgrave, D. & Cho, K. & Oh, A. (eds.), *Advances in Neural Information Processing Systems*, vol. 35. 27730–27744. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.

Shieber, Stuart M. 2004. Introduction. In Shieber, Stuart M. (ed.), *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, 1–13. Cambridge, MA: MIT Press. doi:10.7551/mitpress/6928.003.0002. https://doi.org/10.7551/mitpress/6928.003.0002

Turing, Alan M. 1950. Computing machinery and intelligence. *Mind* LIX(236). 433–460. doi:10.1093/mind/LIX.236.433. https://doi.org/10.1093/mind/LIX.236.433

Wang, Pei. 2019. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence* 10(2). 1–37. doi:10.2478/jagi-2019-0002. https://doi.org/10.2478/jagi-2019-0002

Warwick, Kevin & Shah, Huma. 2016. Can machines think? A report on Turing test experiments at the Royal Society. *Journal of Experimental & Theoretical Artificial Intelligence* 28(6). 989–1007. doi:10.1080/0952813X.2015.1055826. https://doi.org/10.1080/0952813X.2015.1055826