ChatGPT: Not Intelligent

Barry Smith

In our book *Why Machines Will Never Rule the World*,[1] Jobst Landgrebe and I argue that the efforts of many in the artificial intelligence community to create an artificial *general* intelligence (AGI) are doomed to fail. Here "AGI" is defined as referring to a machine that would exhibit cognitive capacities equivalent to, or even surpassing, those of human beings.

Our argument for the impossibility of such a machine has the following form:

1. we analyze the properties of complex systems[2] such as the Earth's weather system or the traffic system of Istanbul,
2. we demonstrate that there are severe limits on our ability to predict mathematically the behaviours of systems of this sort,
3. we show that these limits then determine also the abilities of computers to make such predictions.

---

[1] Jobst Landgrebe and Barry Smith, *Why Machines Will Never Rule the World:. Artificial Intelligence Without Fear*, Abingdon, UK: Routledge, 2022.

[2] We draw here also on Stefan Thurner, Rudolf Hanel and Peter Klimek, *Introduction to the Theory of Complex Systems*, Oxford University Press, 2018. Note that "complexity" as we use it here is distinct from mere complicatedness. As we shall see, there can be very complicated systems which are yet *simple* – in the book we call them "logic systems" – in the sense that their behavior can be predicted by a machine. That this is so is indeed a trivial matter, since the very act of *providing outputs* by ChatGPT is mathematically a case of *predicting those outputs*.

Our conclusion as to the impossibility of AGI now follows from the fact that all organic systems – including the human neurological system – are complex systems in the sense defined in our book. AGI, however, would require computers which are able to predict with high reliability the behaviours of human beings. This ability is indispensable, for example, if a machine is to participate intelligently in human conversations. And it is the lack of this ability which explains the problems we encounter when dealing over the phone with our bank's computers. For the latter still display a miserable level of performance even after 50 years of attempts by AI engineers to simulate human telephone behaviour. Sophisticated competence in communicating with human beings – for example with controllers of armies, suppliers of munitions, and so forth – would be needed, of course, if machines are to rule over the world.

Our arguments imply that computers will always be restricted to the use of algorithms of the sort that are able to predict only the behaviours of simple systems such as laptops or factory assembly lines. For this reason they will always exhibit what is called "narrow AI", which means that they will always fall short of the *general* intelligence of human beings.

Enter ChatGPT

Or so, when our book was published in September 2022, we had claimed.

But then, in November of 2022, a new and revolutionary kind of AI was unleashed onto the world in the form of "ChatGPT". ChatGPT was different. It represented the first example of an AI system that could be accessed easily by normal human beings and provide hours of annoyance-free stimulation. In addition, it was able to provide a range of different kinds of services that can be of value, for example, to commercial organizations. ChatGPT was also, it seemed,

an example of *general* AI, since it could respond, in its way, to prompts on every subject under the sun.[3]

Very soon, however, the growing community of users of ChatGPT became aware of certain unanticipated behaviours, called "hallucinations", on the part of its algorithm. For the latter is apparently statistically primed to avoid phrases like "I don't know" in its outputs. Instead it will oftentimes invent its own answers to questions where it did not know the answer. In short, it would exhibit behaviour which, if committed by a human, would be referred to as "telling lies".[4]

How ChatGPT Works

To understand what is going on here, we must recognize, first, that ChatGPT does not in fact *know* anything. When it gives a response R to a prompt P this is not because it knows that some proposition R is true.

ChatGPT is a set of algorithms in software form, built on the basis of the GPT probabilistic model of the order of the symbol sequences produced, for example, in conversations or in question-and-answer sessions.

When we enter some prompt P, therefore, then what happens is that the ChatGPT algorithm predicts that, given the data that was used as its training set, a given sequence of tokens (roughly: syl-

---

[3] Blaise Agüera y Arcas and Peter Norvig, "Artificial General Intelligence Is Already Here", *Noema Magazine*, October 10, 2023, https://www.noemamag.com /artificial-general-intelligence-is-already-here/.

[4] The conversation displayed here: https://buffalo.box.com/v/ChatGPT-getting-worse provides an example of this phenomenon, illustrating how ChatGPT is of the opinion that there are three philosophers by the name of J. K. (or J. C.) Nyíri. One, a Jozef (dead since 2018), being the father to the next, a Janos, still alive and "known for his work in logic, philosophy of science, and philosophy of language"; the third, a certain "J. Christoph Nyíri", is said to be an emeritus professor of philosophy in Paris and an expert on the topic of heritagization.

lables) R is (roughly) the *next most likely sequence of tokens* given P as starting point.[5]

   We must recognize, second, that ChatGPT functions always only by drawing on the specific (and admittedly very large) body of data upon which it was trained. This implies that ChatGPT, too, is an example of narrow AI. This is because its algorithm does not relate to the real world in which we live and to the many, many, overlapping and ever evolving complex systems out of which it is composed. Rather, it relates to a certain abstract simulacrum of a world, a simulacrum that is exactly specified by the large but finite set of data upon which the algorithm was trained, in something like the way in which the world of a video game is specified by the game's software. For ChatGPT this training data was defined (roughly) by the contents of the internet on some given day in the past.[6]

   The algorithm, which was defined through a process of training on the basis of data available in 2021, is a mathematical function which takes as inputs binary vectors encoding prompts P, and outputs binary vectors encoding responses R. We can think of this function as a very, very long polynomial equation (with some 1.5 billion parameters). Hence the broad reach of topics upon which it can provide responses to prompts. At the same time, however, the mathematical abilities encapsulated in this equation are still very simple – since the equation, like all computable algorithms, needs to be capable of being solved by using only the very simple mathematics of a Turing machine.[7]

---

[5] Note that "predict" here means that the algorithm issues R as output given P as input.

[6] Hence its use of the modifier "As of my last knowledge update in September 2021" to justify its not providing answers to questions pertaining to more recent events.

[7] ChatGPT is in this respect comparable to Google Translate. The latter can be applied to many languages used to cover many, many topics. Yet Google Translate is nonetheless an example of narrow AI. Like ChatGPT it works only for very simple and unchanging worlds, which are defined by the data upon which any given version of the software was trained.

Conclusion

We can now understand why, as the material referenced in footnote 4 demonstrates, ChatGPT yields such peculiar results when questions are raised about persons with names like "J. K. Nyíri", even as it does so well when questions are raised about, say, J. K. Rowling. This is because the internet is rich with data about all things Harry Potter, and the algorithm performs very well when predicting the (next most likely) answers to questions about topics about which it has large amounts of data. It makes such a mess of itself, in contrast, when addressing the topic of a "J. K. Nyíri", because the world of the internet is to such an overwhelming degree an Anglosaxophone, rather than a Magyarophone, world.

We note, finally, that things will not get better, from the point of view of AGI, even if such "hallucination" problems can be solved. If the GPTs of the future can reduce the degree to which they generate hallucinations, it will still remain the case that the world in relation to which each successive release of GPT provides its responses is an unchanging logic-system-defined world. Thus it is not anything like the world in which, over millions of years, human intelligence has evolved.

Barry Smith is Professor in the University at Buffalo. He is the author of more than 700 scientific publications, including 15 authored or edited books, and he is one of the most widely cited living philosophers. E-mail: ifomis@gmail.com.