In Our Own Image: What the Quest for Artificial General Intelligence Can Teach Us About Being Human

JANNA HASTINGS

Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich

School of Medicine, University of St. Gallen

Swiss Institute of Bioinformatics

Centre for Behaviour Change, University College London

Artificial intelligence (AI) has been an inspiration and a goal over the better part of the past century. In the eyes of the public, AI in the form of autonomous and intelligent machines has elicited in equal measure hopes for utopian futures and fears of nightmarish scenarios. Algorithmic progress in automated digital systems has been steady over that time, and such systems underpin many of the digital facilities of our modern world. However, the vast majority of such progress has been dull: infrastructures and algorithms that automate small components of overall pipelines in ways that make workflows slightly more efficient. The anticipated futures of truly intelligent and autonomous agents seemed very far away. Nevertheless, the recent advances in AI technologies have brought the topic of artificial intelligence squarely into the centre of public concern.

In November 2022, OpenAI released ChatGPT (OpenAI 2022), and with it came a new generation of AI tools able to produce plausible-sounding responses in well-formed natural language to prompts across a wide range of different topics and formats. ChatGPT and other large language models have been reported to have broken the longstanding 'Turing test' for the intelligence of an automated system (Biever 2023): they have reportedly managed to fool humans into believing that they are human. But does this really mean, as Turing anticipated, that such language models are intelligent? The answer to this question is not so straightforward to determine, as it depends both on what is meant by intelligence, and the specific formulation of the contexts in which the test is to be applied. There is however no question that language models do not approach human-level intelligence, for a multitude of reasons. They can process language, but they cannot reason abstractly or do mathematics (Frieder et al. 2023; Mitchell 2023), nor do they have common sense (Bian et al. 2023), nor can they act in the world, nor are they conscious, nor do they have any independent identity as individuals. Nevertheless, their linguistic capabilities are uncanny and impressive. For many members of the public, and for many researchers, the achievements of this new paradigm of generative AI and the rapid pace at which such achievements have been realised strongly suggest that human-level artificial intelligence (artificial general intelligence, AGI) must be imminent. This perspective is echoed by prominent technology researchers and human scientists (Bubeck et al. 2023; Harari 2023). These technological advances and their wide reach and impact make it ever more urgent and pressing to understand the implications of these technologies and their true potentials—what can they be expected to be capable of in the future?

In August 2022, only a few months before ChatGPT was released, Barry Smith, well-known contemporary philosopher, together with Jobst Landgrebe, artificial intelligence entrepreneur, published a book entitled *Why Machines will Never Rule the World: Artificial Intelligence without Fear* (Landgrebe and Smith 2022). In this important, dense and far-reaching work, Landgrebe and Smith argue from the mathematical theory of complex systems, and a sophisticated analysis of the capabilities of human intelligence, that AGI—at the level of human intelligence—will *never* be possible. In broad outline, Landgrebe and Smith claim that intelligent human behaviour is a facility that develops from a complex world and for a complex world; that the complexities of the real world and our behaviour in it can never be fully specified or simulated mathematically; that digital systems can only execute what can be simulated mathematically; and that therefore human-level intelligence will never be manifested by a digital system. Their central argument draws from a nuanced and interdisciplinary depiction of human intelligence as an adaptive complex trait that enables appropriate dynamic and immediate responses to a complex and varying environment, fundamentally outside the realm of what can be adequately modelled mathematically. From this, they argue, it follows directly that intelligence cannot be adequately simulated in a digital system, since everything that executes in a digital computer is a mathematical function, regardless of whether it has been directly specified programmatically or learned from data. If Landgrebe and Smith are correct, then the appearance of seeming intelligence in systems such as ChatGPT is only an illusion, and recent advances in AI technologies are bringing us no closer to human-level AGI, despite their impressive capabilities and the large-scale hype surrounding them.

Landgrebe and Smith succeed in bringing to the forefront of our minds the enormous complexity of the being-in-the-world that constitutes human existence. Their deep interdisciplinary treatment reminds us that our research, from all of our different, often fragmented, disciplinary perspectives, grapples with only small parts of the whole that we are, as complex and multifaceted agents acting in complex societies within a complex world. They also remind us that it is very difficult to discuss the capabilities and limitations of artificial intelligence technologies, as to do justice to the topic requires the intersection of transdisciplinary knowledge across domains—not only from computer science and psychology, but also from physics and mathematics, biology, the social sciences, and philosophy.

The present collection of articles is in response to and inspired by Landgrebe and Smith. It provides reflections from a variety of perspectives and disciplines on the possibility of AGI, as well as on the transformations that AI technology is already bringing even in its current incarnation.

The central question posed by Landgrebe and Smith is the question of whether AGI is possible. Their conclusion is that digital systems can only approximate, but not fully manifest, intelligent behaviour, due to inherent limitations of digital systems. In the first of our collection of articles, Rapaport (2023) takes aim at the idea that the limitations of current technology define the capabilities of future technologies. The advent of modern neural networks to their current level of capabilities depended on a complex set of interweaved innovations in hardware such as GPU processors, software such as new optimisation algorithms, and the availability of large-scale data offered by the development of the World Wide Web. The contribution asks how likely is it that we can anticipate the shape of future computational innovations in the way that Landgrebe and Smith claim? In addition, Rapaport asks, is it indeed necessary that any computational intelligence that is digitally achieved offers a perfect simulation of human intelligence for it to manifest sufficient capabilities to be considered genuinely intelligent? In effect, this contribution takes aim at the claim that approximations of intelligence will never reach the level of capability that humans have.

There is scope to dispute the claims of Landgrebe and Smith from the other direction—are we, as intelligent beings, really operating at a completely different level to the operation of digital systems? In our next contribution in the collection, Simon (2023), takes a deeper dive into the cognitive capabilities of humans, in order to provide a challenge to Landgrebe and Smith's view of human intelligence as non-computational. The key question here is what it means to be computational, and therefore, whether we can indeed claim that computers and other digital systems are doing something different in fundamental nature from what humans are doing.

The definition of AI depends not only on what it means to be intelligent, but also on what it means to be artificial. In our next contribution, Schulz and Hastings (2023) highlight exactly this point, asking in effect to what extent all possible paths towards the creation of AGI are covered by the argument presented in Landgrebe and Smith. In particular, they ask to what extent an artificial system has to be completely engineered in the way that Landgrebe and Smith outline, or whether there would be ways for artificial systems to be designed such that they mix and intermingle with the world and thereby become subject to the types of contextual and environmental influences that have given rise to human intelligence.

Being and acting in the world is a key capability that separates human intelligence from artificial intelligence. In our next contribution, Martinelli (2023) is broadly in agreement with the arguments and conclusions presented in Landgrebe and Smith, but offers an additional argument which they do not present. This argument draws attention to the ontological difference between the particularity of an individual organism's interactions with the world and the general, averaged, mediated and non-particular interactions that AI has with the world.

A different type of challenge to Landgrebe and Smith is posed by the next series of articles in the collection. These take as their starting point the consideration that even if AGI is indeed impossible, as Landgrebe and Smith argue, there are certainly still going to be many practical implications for our modern world posed by the development of systems with an ever-increasing series of approximate AI capabilities. In the next contribution in our collection, Fjelland (2023) raises the possibility that Landgrebe and Smith may be too confident in their assertion that AI without the spectre of AGI should inspire no fear. This contribution highlights that AI without AGI is already transforming our world, and we are already witnessing the consequences of this transformation.

Continuing this discussion, West (2023) in our next contribution frames the quest for AGI from a social and psychological perspective, and asks important questions about who owns and drives the development of AI technologies and for whom they are developed. Who stands to benefit, and who is at risk? This question is among the most urgent that society as a whole must grapple with, given the transformative capabilities that even approximate AI technologies can possess, and the power, economic and otherwise, that these technologies bring to their bearers.

A question that has been challenging for the entire history of the development of AI technologies and which has now become urgent with the modern language models, is how we can operationally evaluate the capabilities of such systems. The tests that are used to evaluate the capabilities of humans do not translate straightforwardly to evaluating the capabilities of automated systems that are trained on human-produced content. In our next contribution, Krinkin (2023), a key theme is the need for new forms of practical test for computational intelligence that go beyond the Turing test which is all too easily "solved" by the mimicry of modern AI systems. The suggestion is that new forms of test would involve cooperation on a shared common task in the world.

In the final section of our collection, we conclude with some articles that look specifically at the implications of the modern generative AI technology that underlies both language models and other transformative technologies such as image generation. In this section, the contributions address topics or questions that are not explicitly tackled in Landgrebe and Smith.

Sedlakova (2023) looks specifically at the use of AI in conversation for psychotherapeutic purposes, and discusses the broader practical and ethical implications of the use of the technology in that context. Conversation is at the heart of human intersubjective experience, but what does it mean to create and construct intersubjectivity in conversation with systems that have no subjectivity themselves? This question has immediate practical relevance, as mental health and psychotherapy is already a large application domain for digital health applications.

In the final contribution of the collection, Hedblom (2023) explores to what extent modern generative AI image generation tools can be said to be creative. Image generation technology, and related technologies such as audio generation and video generation, are changing the way that art itself is made. The creation of art may be one of the most essential and yet mysterious features of being human, not necessary for our sur-

vival in any crude evolutionary sense, yet an inseparable part of the deeper and higher aspects of human experience. Image generation therefore provides a different sort of window into what AI is and what it can and cannot do.

The collection is then rounded off with a response by Landgrebe and Smith to each of the individual contributions (Landgrebe and Smith 2023), reflecting their perspectives on the wide range of ideas contained in the individual contributions as well as, ultimately, a re-affirmation of their view that recent developments in AI in no way shake their central thesis.

Landgrebe and Smith pose a strong challenge to the quest for AGI, and by so doing they ask us to become more aware of what it means to be a human navigating in a complex world, as we try to create systems in our own image. Indeed, many of the challenges that AI cannot solve today are those that form a part of the deeper and less rational aspects of being human. It would be fair to observe at this point, as also observed in Landgrebe and Smith, that for many aspects of human functioning we fundamentally do not know of ourselves how we operate as we do. Regardless of whether AGI is possible or not, the quest to reach it demands of us that we study ourselves more deeply, that we become more able to reflect upon and characterise the richer and fuller aspects of experience in the world that go beyond what is superficial and what can be trivially captured in data. And in so doing, we may yet discover that we are much more complex than we thought we were.

REFERENCES

Bian, N., Han, X., Sun, L., Lin, H., Lu, Y., and He, B. 2023. *ChatGPT is a Knowledgeable but Inexperienced Solver: An Investigation of Commonsense Problem in Large Language Models*. https://doi.org/10.48550/arXiv.2303.16421.

Biever, C. 2023. ChatGPT broke the Turing test—The race is on for new ways to assess AI. *Nature*, 619(7971):686-689. https://doi.org/10.1038/d41586-023-02361-7.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., and Zhang, Y. 2023. *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. https://doi.org/10.48550/arXiv.2303.12712.

Fjelland, R. 2023. Computers will not acquire general intelligence, but may still rule the world. *Cosmos + Taxis* 12:5+6.

Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A., and Berner, J. 2023. *Mathematical Capabilities of ChatGPT* https://doi.org/10.48550/arXiv.2301.13867.

Harari, Y. N. 2023. *AI and the future of humanity | Yuval Noah Harari at the Frontiers Forum*, May 14. https://www.youtube.com/watch?v=LWiM-LuRe6w.

Hedblom, M. M. 2023. Every dog has its day: An in-depth analysis of the creative artistic ability of generative AI. *Cosmos + Taxis* 12:5+6.

Krinkin, K. 2023. Back to evolutionary intelligence. *Cosmos + Taxis* 12:5+6.

Landgrebe, J., and Smith, B. 2022. *Why Machines Will Never Rule the World: AI without Fear*. New York and Abingdon: Routledge.

_____. 2023. Intelligence. And what computers still can't do. *Cosmos + Taxis* 12:5+6.

Martinelli, E. 2023. Complexity and Particularity: An Argument for the Impossibility of Artificial Intelligence. *Cosmos + Taxis* 12:5+6.

Mitchell, M. 2023. How do we know how smart AI systems are? *Science*, 381(6654). https://doi.org/10.1126/science.adj5957.

OpenAI. 2022. *ChatGPT*. https://chat.openai.com.

Rapaport, W. J. 2023. Is Artificial General Intelligence Impossible? *Cosmos + Taxis* 12:5+6.

Schulz, S., and Hastings, J. 2023. What is a machine? Exploring the meaning of 'artificial' in 'artificial intelligence'. *Cosmos + Taxis* 12:5+6.

Sedlakova, J. 2023. Conversational AI for Psychotherapy and Its Role in the Space of Reason. *Cosmos + Taxis* 12:5+6.

Simon, J. A. 2023. Is Intelligence Non-Computational Dynamical Coupling? *Cosmos + Taxis* 12:5+6.

West, R. 2023. Semi-autonomous Godlike Artificial Intelligence (SAGAI) is conceivable but how far will it resemble Kali or Thor? *Cosmos + Taxis* 12:5+6.