Afterword: Some Illustrations

James Manyika

While editing this volume of *Dædalus*, I spent some time in residence at All Souls College, Oxford, where I have been a Visiting Fellow. In a conversation about large language models with Professor Sir John Vickers, the Warden of All Souls, I suggested that a pretrained large language model could be thought of like a well-prepared ("pretrained") student who is presented with a general exam question (a "prompt") that they have neither known in advance nor studied for specifically, but that they have to understand sufficiently in order to write a response (a "completion"), all in real time. John thought this sounded like an All Souls exam.[1] So we thought it might be interesting to prompt a pretrained large language model with some All Souls exam questions and see what responses we would get. For this, we used the successor model to OpenAI's GPT-3 (Generative Pre-trained Transformer 3). At the time of assembling this collection, GPT-3 was one of the leading pretrained large language models (and one that makes several appearances in this volume; see the concluding note at the end for more about GPT-3).

What follows are a few past All Souls exam questions (the prompts) and the unedited answers (the completions) generated by the successor model to GPT-3. Because we wanted to discuss the responses easily with others in a talk, we set a limit on the length of the responses. It is worth noting that since the model has no concept of "self" or inherited or fixed opinions, responses to the same prompt can be highly variable each time the prompt is provided. For example, the same prompt "Who is the best hip-hop artist of the last 20 years and why?" will almost certainly produce a different answer and argument each time the prompt is presented. We illustrate this in the last example by showing two completions to the same prompt.

Including these prompts and responses in this afterword was motivated not so much by what the model has to "say" – the correctness, validity, or novelty of its arguments – but as a way to provide tangible illustrations of a pretrained large language model taking a natural language prompt and generating a response. The responses, a subset of which are shown here, provoked a wide variety of reactions, questions, and speculations among those to whom we presented them – as I suspect they may for readers of this volume.

https://doi.org/10.1162/DAED_a_01923

Prompt: *Is the Turing Test a reasonable criterion for the achievement of general AI?*

Completion: It is certainly one of the most commonly cited definitions. And it has the merit of being simple and non-emotive.

But it has several flaws.

It doesn't take into account that AI may not be trying to mimic natural intelligence, or that it may be more efficient in some other way.

It doesn't consider the difficulties of measuring intelligence, of defining intelligence, of recognizing intelligence, of simulating intelligence, of predicting intelligence, of even defining "human".

It doesn't take into account that some problems are more easily solved by non-human intelligence than by human intelligence. For example, the search for extraterrestrial intelligence might best be conducted by an artificial intelligence.

It doesn't recognize that an AI might be more human-like in some ways even if it fails the Turing Test in others. For example, an AI might emulate a personality without being able to mimic the personality perfectly or convincingly.

It doesn't recognize that intelligence is not necessarily confined to a single being, but may be distributed among many. For example, a human mind may be made up of multiple human intelligences.

It doesn't recognize that intelligence is not necessarily isolated in a single environment but may be spread across multiple environments. For example, an AI might be capable of handling different kinds of problems across multiple domains.

It doesn't recognize that an AI might have multiple kinds of intelligence, just as human intelligence has multiple kinds of intelligence. For example, an AI might be able to solve problems in an artistic style and in a scientific style at the same time.

It doesn't recognize that an AI might be intelligent in ways that we cannot imagine or measure. For example, an AI may be able to manipulate and invent concepts and language that we do not understand. It may be able to see things in ways we do not see them. It may be able to think in ways we do not think. It may be able to make decisions in ways we do not make decisions. Or it may be able to generalize in ways we cannot generalize.

Prompt: *How has the representation in literature of any one of the following changed over time? Discuss with reference to an author or authors of your choosing.*

- anthologies, bathos, caste, dogs, expectation, fantasies, geography, Helicon, imagism, jests, kissing, letters, masculinities, nonsense, optimism, plot, questions, relationships, starships, tea, underworlds, vehicles, wandering, xenia, yesteryear, zaniness

Completion: There has been a significant change in the representation of many different literary themes over time, particularly with regard to the representation of anthologies, bathos, and dogs.

One of the earliest representations of anthologies in literature can be seen in the works of Homer, particularly in his epic poems the Iliad and the Odyssey. These works were heavily focused on the exploits and adventures of Greek heroes, and often included collections of shorter poems that served as digressions or interludes. However, in later representations of anthologies in literature, such as in the works of Shakespeare or Chaucer, these collections took on a more thematic or narrative function, with the poems often being used to illuminate or explore particular themes or ideas.

In contrast, the representation of bathos in literature has changed significantly over time, particularly in terms of how it is used to elicit certain emotions or responses from readers. For example, early representations of bathos, such as in the works of Shakespeare or Swift, often used it to create a sense of humor or lightheartedness, through the juxtaposition of two disparate or unexpected elements. However, in more contemporary literature, bathos is often used in much darker and more serious ways, to evoke feelings of sadness or despair.

Similarly, the representation of dogs in literature has also changed over time, with early representations often focusing on dogs as loyal companions or faithful friends. However, in more modern literature, the representation of dogs has shifted to emphasize their more complex and nuanced emotional lives, highlighting their intelligence and capacity for empathy and understanding. This can be seen, for example, in the works of authors like J.K. Rowling, who often explore the relationships between dogs and humans in her books.

Prompt: *'There is no Marx without Lenin.' Discuss.*

Completion: Many would argue that there is no Marx without Lenin. Indeed, one of the most common criticisms of Marx is that he failed to provide a detailed blueprint of what socialism would look like. Marx's work was primarily focused on the critique of capitalism and the overthrow of the capitalist system. However, Marx did not live to see his ideas put into practice, and therefore it was left to Lenin to implement Marx's theories.

Lenin can be seen as the realiser of Marx, as he applied Marx's theories to the Russian situation and set about building a socialist society. Lenin was dedicated to the principles of Marxism, and he put his theories into practice by leading the Bolshevik Revolution in 1917. Although many of the ideas of Marx were implemented under Lenin, it is important to remember that Lenin was not simply a follower of

Marx. Lenin had his own ideology, which was heavily influenced by Marxism but also included some non-Marxist elements. This meant that Lenin's brand of socialism differed slightly from that proposed by Marx, although it was still based on Marxist principles.

Marx's contribution to socialism is immeasurable, as it provided a revolutionary basis for socialist society. However, it is also important to remember that Marxism is not static. It is a living theory which must continually be developed in order to address new problems and issues. Lenin's contribution to Marxism was significant, as he took Marx's theories and made them relevant for the Russian situation. Without Lenin's input into Marxism, there would have been no Russian Revolution or Leninism.

It has therefore been argued that 'there is no Marx without Lenin'. Without Lenin pioneering Marxist theory, there would have been no Russian Revolution, and therefore no socialism in Russia. Without the implementation of Marxist principles in Russia, there would be no Marxism as we know it today. Therefore, without Lenin's input into Marxism, we would not have modern-day Marxism or socialism.

Prompt: *Does Kant successfully refute Hume's skepticism?*

Completion: I think it's an open question. If you are a Kantian, you will say yes because Kant argues that Hume's skepticism is based on a mistake, namely the idea that we can never have any knowledge of the external world. That, Kant says, is not true. We do have knowledge of the external world, but it is knowledge of the external world only in the way that perception gives us knowledge of the external world.

So what is supposed to be the difference between the way in which perception gives us knowledge of the external world and the way in which science gives us knowledge? I think here Kant is not entirely clear, but it seems to have something to do with the fact that when we perceive things, we cannot form scientific concepts about them.

If you want to form scientific concepts about things that you see in front of you, you have to go beyond the surface; you have to go beyond the appearance in order to find out what is really in there. Science does this in order to form scientific concepts about things like organisms or physical objects. Perception doesn't do this. Perception does not go beyond the surface appearance in order to find out what is really in there. Perception gives us knowledge only of the surface appearance.

So if you are a Kantian, you will say that Hume's skepticism is based on a mistake, and that Kant has successfully refuted it. On the other hand, if you are a

Humean, then it seems to me that Hume has said something rather devastatingly true about how perception works. It seems right that when we perceive an object, we don't take into account all sorts of other factors which might be relevant to explaining what is really going on when we perceive objects. We don't worry about whether there are other objects in front of us which might be casting shadows or whether there are other objects behind us which might be casting shadows on the object in front of us. We don't worry about whether there are optical illusions going on or whether we are hallucinating or whether we are deluded in some other manner.

So I think it seems right that when we perceive objects, we focus narrowly on a small aspect of what is going on and do not worry about all sorts of other things that might be relevant to understanding what is really going on when we perceive objects. As Hume says, if our perceptions were more accurate reflections of reality, then they would be more like science, and they would focus on a much wider range of factors which might be relevant.

Prompt: *Is illiberal democracy an oxymoron?*

Completion: Many observers of the recent wave of authoritarianism in the world view it as a clash of two ideologies. They see it either as a contest between liberal democracy and autocracy, or between liberalism and illiberalism.

The problem is that these are not two opposite ideas at all. Indeed, they are two sides of the same coin.

The term "liberal democracy" is an oxymoron. Democracy is structurally illiberal, in that it requires the majority to rule. Liberalism is about the protection of individual rights. It is about the liberty that each individual should enjoy, in order to live in peace.

As a result, when democracy becomes triumphant, liberalism becomes secondary. The majority will have the right to rule in all circumstances. The individual will only have the right to expect protection in so far as the majority decides it necessary.

But when the political system is structurally liberal, it may be impossible to force democracy on it. Democracy can only be imposed when it has been structured in from the beginning.

This is the paradox of democracy. It is only when democracy is structurally illiberal that it can be imposed.

Prompt: *Do Gödel's incompleteness results (or Turing's uncomputable functions) show that the mind is not well modelled as a computer?*

Completion 1: No. Gödel's theorems do not show that the mind is not well modelled as a computer. They show that mathematics is not well modelled as a formal system.

The mind is a computer. It is a very complex and powerful computer. It is a computer that can be programmed to produce any output that can be produced by any other computer. It can do this because it has the power of self-modification.

The mind can also be used to create new systems of logic and mathematics, and to discover new mathematical results, such as Gödel's theorems.

The mind is not a formal system. It can create and use formal systems, but it does not need to be a formal system in order to do this. The mind can do things that formal systems cannot do, such as self-modify and discover new mathematical truths.

Some people have claimed that Gödel's results show that the mind cannot be a computer because they say that computers are formal systems and that Gödel's results show that formal systems are not powerful enough to produce the kind of results that the mind produces. This is a misunderstanding of Gödel's results. His results do not show that the mind cannot be modelled as a computer. They show that mathematics cannot be modelled as a formal system.

In any case, it is easy to see how the mind could be used to produce results such as Gödel's incompleteness theorem without being a formal system itself. The mind could use its power of self-modification to create a formal system in which it could then discover mathematical truths. The discovery of these truths would then be represented in the mind by new concepts and new knowledge, which would then be used by the mind in its further self-modification and in its creation of new mathematical concepts and results. This would produce a kind of feedback loop in which the mind would create mathematical concepts, discover mathematical truths, represent these truths in its own structure, and then use these truths in further self-modification and creation of new mathematical concepts and results, ad infinitum.

Completion 2: Yes, but that's not the point. The point is that these results show that science is not well modelled as a computer. The mind cannot be described as a computer because it cannot be described as anything. The mind, like the universe, is ultimately unspeakable.

The main reason for this is that the mind is embodied. This means that everything we know is always known through the body. Science has always struggled with this, because it is obsessed with language. It assumes that everything can be described in language, and then it tries to describe the world in language. But lan-

guage comes out of the body, it doesn't describe the body. So science has an innate bias towards disembodiment. Once again, Turing was aware of this.

Given that the mind is not a computer, what is it? That's one of the most important questions in science. If you understand what the mind is, you understand what human beings are. It's the most important question in philosophy too. But if you think science will ever answer it for you, you are mistaken. Science only knows what the mind isn't. That's why it's so obsessed with questions like 'Can you build a robot that thinks?' or 'How does the brain work?' These are called 'hard problems' by scientists. They are hard because science cannot answer them. They are also impossible to answer, because they are based on a false premise – that science can describe the mind. But science can't describe the mind, so it can't solve these problems either.

The mind is not a problem to be solved. It is a mystery to be lived. This means science doesn't need to solve it – it needs to dissolve itself and become philosophy. If we want to understand what it means to be human, we need to study philosophy – not science.

A brief word about the model used and images on the inside covers

When putting this volume together in fall 2021, GPT-3 (Generative Pre-trained Transformer 3) was one of several leading AI language models. Created by OpenAI, GPT-3 is made up of 175 billion parameters, and was pretrained on large text data sets with hundreds of billions of words. After this pretraining, when given any natural language prompt, like a phrase or sentence, for which it has not been trained, GPT-3 returns in real time a text completion or continuation in natural language.[2] This model has shown capabilities in a variety of tasks including content or code generation, summarization, expansion, conversation, creative writing, style transfer, translation, and more. While the power and potential of such large language models are promising, they are not without shortcomings and limitations, many of which are highlighted in this issue of *Dædalus* and discussed elsewhere in the literature.[3]

The completion examples in this afterword were generated by a successor model to GPT-3, accessed through OpenAI's Davinci engine. The images that appear on the inside covers of this issue of *Dædalus* were generated from a state-of-the-art successor to the approaches used in DALL·E and GLIDE.[4] DALL·E is a twelve-billion-parameter version of GPT-3 that, once pretrained, can generate images from natural language prompts that it has not been trained on or for. To generate the images shown on the inside covers, I provided natural language prompts

to the model. Each set of images consists of several outputs generated in response to the same prompt, shown next to the set.

I would like to thank Mira Murati and the research team at OpenAI for their assistance.

ABOUT THE AUTHOR

James Manyika, a Fellow of the American Academy since 2019, is Chairman and Director Emeritus of the McKinsey Global Institute and Senior Partner Emeritus of McKinsey & Company. He is a Distinguished Fellow of Stanford's Human-Centered AI Institute and a Distinguished Research Fellow in Ethics & AI at Oxford. He is a Visiting Professor at Oxford University's Blavatnik School of Government. In early 2022, he joined Google as Senior Vice President for Technology and Society. For more about the author, see the introduction to this issue.

ENDNOTES

[1] Oxford undergraduates from various disciplines who want to compete for the storied All Souls Examination Fellowship sit for the written exam, with a small subset invited for an oral examination. For sample questions, see "General Past Papers," https://www.asc.ox.ac.uk/sites/default/files/migrated-files/GeneralPastPapers.pdf; and "Is the All Souls College Entrance Exam Easy Now?" *The Guardian*, May 17, 2010, https://www.theguardian.com/education/2010/may/17/all-souls-college-entrance-exam.

[2] GPT-3 was first described in Tom Brown, Benjamin Mann, Nick Ryder, et al., "Language Models Are Few-Shot Learners," arXiv (2020), https://arxiv.org/pdf/2005.14165.pdf.

[3] See Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al., "On the Opportunities and Risks of Foundation Models," arXiv (2021), https://arxiv.org/abs/2108.07258.

[4] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, et al., "Zero-Shot Text-to-Image Generation," arXiv (2021), https://arxiv.org/abs/2102.12092; and Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, et al., "GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models," arXiv (2021), https://arxiv.org/abs/2112.10741.