LETTERS on LIBERTY

AI: SEPARATING MAN FROM MACHINE

Sandy Starr

LETTERS on LIBERTY

Welcome to *Letters on Liberty* from the Academy of Ideas. *Letters on Liberty* is a modest attempt to reinvigorate the public sphere and argue for a freer society.

academyofideas.org.uk/letters

academy of ideas

Since its foundation in 2000, the Academy of Ideas has hosted thousands of public debates, festivals, forums and salons where people from all walks of life come together to debate often-controversial topics and to challenge contemporary knee-jerk orthodoxies. We always hold on to one defining principle: free speech allowed.

academyofideas.org.uk

What are Letters on Liberty?

It's not always easy to defend freedom. Public life may have been locked down recently, but it has been in bad health for some time.

Open debate has been suffocated by today's censorious climate and there is little cultural support for freedom as a foundational value. What we need is rowdy, good-natured disagreement and people prepared to experiment with what freedom might mean today.

We stand on the shoulders of giants, but we shouldn't be complacent. We can't simply rely on the thinkers of the past to work out what liberty means today, and how to argue for it.

Drawing on the tradition of radical pamphlets from the seventeenth century onwards - designed to be argued over in the pub as much as parliament - *Letters on Liberty* promises to make you think twice. Each *Letter* stakes a claim for how to forge a freer society in the here and now.

We hope that, armed with these *Letters*, you take on the challenge of fighting for liberty.

Academy of Ideas team

AI: SEPARATING MAN FROM MACHINE

The author of this *Letter on Liberty* is a human, and he assumes that the same is true of you. What guarantee do either of us have?

This question is reminiscent of the Turing test - the 'imitation game', famously proposed by computing pioneer Alan Turing, which assesses a machine's ability to imitate a human convincingly.[i] The question has been given renewed urgency by today's 'generative' artificial intelligence (AI).

AI encompasses a wide range of technologies, developed since the 1950s, that supposedly emulate things done by human brains and/or human minds (delete according to one's philosophy). The AI tools that currently dominate the headlines are called 'generative' because they generate seemingly unique, bespoke creations - text, images, code - in response to 'prompts' submitted by people.

The results can be spectacular. There was controversy when first prize in the 'digital arts' category at the 2022 Colorado State Fair fine art competition was awarded to *Théâtre d'Opéra Spatial*, a piece created via generative AI by gaming company CEO Jason Allen. My imagination was fired by this (pretentiously titled)

image of robed, vaguely humanoid forms, two of whom - one sporting a bustle and perhaps a spindly appendage like the legs of Dalí's Elephants, another echoing the attitude of Tenniel's Red Queen - are looking at, or through, a great disc or globe. I found the image compelling regardless of how, by whom or by what it had been created.

But we can't always be so relaxed about the matter. When we are reading an article, or marking homework, or exchanging spoken or written words, most of us want to know whether the other party is a human being. Short of being in the same physical space, in an age of generative AI and so-called 'deepfakes' we can ultimately only rely on two things to ascertain one another's humanity: the quality of our attention and the exercise of our judgment.

Are these enough?

Accustomed to wonders

The late, great American film critic Roger Ebert once wrote a favourable review of a *Star Wars* film, now considered by many to be the nadir of that series. Ebert was no slouch when it came to exercising judgment, but he was also refreshingly uncynical. He

thought that, in this instance, the special effects were so groundbreaking that it was incumbent upon the critic to pause and take stock. 'How quickly do we grow accustomed to wonders', he wrote.[ii]

Today's commentary on generative AI may date as badly as that *Star Wars* film, but we should still take a moment to do an Ebert. We should acknowledge the ingenuity of tools that can take prompts in natural human language, and generate sophisticated material in response.

Nothing can come of nothing, and generative AI comes from a long and rich history.

At the time of writing, material generated in this way can take the form of written or spoken text or programming code (from ChatGPT or Bard) or static images (from Stable Diffusion, Midjourney or DALL-E). Other possibilities coming down the track involve moving images, music, facial recognition and synthesis, voice recognition and synthesis and the ability to submit images or sounds (rather than words) as prompts.

Two breakthroughs in the field of machine learning have ushered in the latest possibilities. The first, made in 2015 by researchers at Stanford and Berkeley, was

the invention of 'diffusion' models.[iii] Taking inspiration from the phenomenon of diffusion as understood in physics, these models begin by training machines to mess images up, turning clear images into incoherent noise.

The idea is that if the machine is trained in creating nothing from something, it can then be trained to reverse the process and create something from nothing. Or rather - forestalling King Lear's objection that 'nothing can come of nothing'[iv] - to create something from the germ of a written prompt, and from an enormous dataset of images on which the machine has previously been trained.

Images generated in this way cannot be dismissed as simple regurgitations of existing material. They are more novel and interesting than that. They are also more eerie - even including photorealistic images of people who have never existed.

The second breakthrough involves 'large language models', which are ways of applying machine learning to vast quantities of text. In 2017, researchers at Google invented a 'transformer' architecture for these models.[v] Compared with earlier approaches, this entailed a more exclusive focus on - and a more flexible means of - emulating the human process of attention. Transformer architecture accounts for the

'T' in 'ChatGPT', a tool whose creators - OpenAI - proceeded to steal a march on Google.

The results of these breakthroughs speak for themselves (sometimes literally), and it can often feel as though this technology has sprung up from nowhere. But Lear's objection, disastrously wrong when applied to his youngest daughter, is correct in this context. Nothing can come of nothing, and generative AI comes from a long and rich history.

Poetical science

Generative AI is the culmination of an idea that humanity has been mulling over for centuries - the idea of presenting something *to* some magical or mysterious creature, deity or device, and then getting a transformed or newly created thing back *from* it. Nowadays, we call the thing submitted the 'input' (or in the case of generative AI, the prompt) and the thing given back the 'output'.

We can trace this idea back as far as we like. Even today, theorists of computing conduct thought experiments involving so-called 'oracle machines', borrowing a term that once described the prophets of antiquity. But if we want to understand when the

input/output idea moved from magic to machines, the best place to start is probably the year Victoria inherited the British throne - 1837.

Why 1837? For one thing, it was the year when Lejeune Dirichlet - a German mathematician with an incongruously French name - formalised the modern definition of a 'function', something that takes in a numerical value as its input and spits out a numerical value as its output.[vi] A function won't write an article for you, make a pretty picture or do your homework, but it's a first step on the long road towards making a machine that might do such things.

1837 is also the year when the English polymath Charles Babbage first described his Analytical Engine, the precursor to the modern electronic computer.[vii] Building on his earlier Difference Engine (a mechanical calculator) and on programmable looms devised by various French inventors (eventually patented by and named after Joseph Marie Jacquard), Babbage set about creating a more general purpose device.

Babbage's mathematician colleague Ada Lovelace elaborated on his ideas and came up with what was arguably the world's first ever *computer program* - something that takes in a coded command as its input, and then gets a device to perform a task or series of

tasks as its output. Strictly speaking, Lovelace created the 'execution trace' of a program rather than a program *per se*, but the idea of computer programming is implicit in what she wrote.[viii]

Lovelace coined the beautifully resonant phrase 'poetical science' to describe what she wanted to achieve.[ix] Her upbringing had been defined by the wish of her mathematically minded mother to turn her into a level-headed logical thinker, rather than an errant poet like her father, the (in)famous Lord Byron. But Lovelace refused to accept that poetry and science were antithetical. Her more integrated outlook is apparent in her statement that 'the Analytical Engine *weaves algebraical patterns* just as the Jacquard loom weaves flowers and leaves'.

Lovelace had a nuanced view of what an Analytical Engine (or its successor) might achieve. On the one hand, she thought it might be used to create art - specifically, to 'compose elaborate and scientific pieces of music'. On the other hand, she cautioned that the Analytical Engine was not a creator of entirely original work. She said that it 'has no pretensions whatever to *originate* anything', and that 'its province is to assist us in making *available* what we are already acquainted with'.[x]

These statements from 180 years ago do a remarkably good job of characterising contrasting views of today's generative AI. It is indeed fitting that two of the large language models in the GPT-3 family are named 'Ada' and 'Babbage'.

A random element

It would take a century for another English polymath - Alan Turing - to create a rigorous model for how computers and computer programs work. But something important happened in the hundred years that separated the invention of the computer from the full realisation of the computer - the invention of Markov chains.

If we want to assert the existence of free will, we cannot rely on mathematics to make the case for us.

Markov chains offer a way of drawing probabilistic connections and dependencies between two or more distinct things. When one does this, and then steps back to consider the resulting picture, what is revealed is a subtle interplay of predictability and randomness. This sort of interplay is at the heart of modern

generative AI. This was anticipated by Turing when he argued that 'it is probably wise to include a random element in a learning machine'.[xi]

Disconcertingly, the Russian mathematician Andrey Andreyevich Markov Sr, after whom Markov chains are named, originally invented them in around 1907 to defeat an argument for the existence of human free will.[xii] One might conclude from this that Markov should be given short shrift in a *Letter on Liberty*, but, if anything, he did the cause of liberty a favour.

Markov's specific quarrel was with rival theoreticians who held that the existence of free will could be proved *mathematically*, an argument that was undone by the properties of Markov chains.[xiii] Markov demonstrated that if we want to assert the existence of free will, we cannot rely on mathematics to make the case for us.

Not content with using his chains to win that argument, Markov used them again in 1913 to analyse the first 20,000 (typographical) characters of Alexander Pushkin's great verse novel *Eugene Onegin*. Markov created a chain that captured the likelihood of a consonant being followed by a vowel, or a vowel being followed by a consonant, in the writing of Pushkin.[xiv] This is a very literal example of someone pursuing Lovelace's 'poetical science'.

A subsequent pioneer of information theory - Claude Shannon - realised that Markov-like methods could be used for generative as well as analytical purposes, to 'approximate to a natural language by means of a series of simple artificial languages'.[xv] But Markov's way of studying Pushkin already begins to resemble, in very rudimentary form, what ChatGPT is doing under the hood if you ask it to write in the style of Pushkin (or whomever).

If generative AI unsettles or deceives us, this may suggest that our attention and our judgment were already wanting

Markov applied his methods painstakingly by hand, using just two parameters - consonant-to-vowel probability and vowel-to-consonant probability. ChatGPT, by contrast, uses rapid electronic computing and billions (as of GPT-4, reportedly trillions) of parameters. When digital necromancy becomes this high-powered, it starts producing a rather more coherent virtual zombie Pushkin (alas, the genuine article remains stubbornly unproductive in his grave in Pskóvskaya Óblast).

Passing the test

We no longer need a brilliantly obsessive oddball like Markov to turn poetry into data, or to attempt the alchemical reverse. Widely available generative AI can now do this for us.

Whether the results satisfy AI theorist Margaret Boden's much-cited definition of 'creativity' as 'the ability to generate novel, and valuable, ideas'[xvi] - and more to the point, who or what deserves the credit if this standard is met - are rich topics for debate.

If we are afraid of the threat generative AI could pose to our freedom, perhaps our belief in and defence of liberty need to be revitalised.

But what is increasingly obvious is that the Turing test is in truth more an assessment of *us* than it is of our technology. If we lose the capacity to distinguish ourselves and one another from machines, then in some sense it is we who have failed a test, rather than our machines that have passed one.

If generative AI unsettles or deceives us, this may suggest that our attention and our judgment were

already wanting - that we were content to behave like machines, or to treat other people as though *they* were machines, before it became feasible that we were literally talking to machines. If aspects of our behaviour, communications and creations can now be emulated by machines, then perhaps we should take this as encouragement to behave, communicate and create differently. If we are afraid of the threat generative AI could pose to our freedom, perhaps our belief in and defence of liberty need to be revitalised.

What is generative AI trying to tell us? At one level, nothing. It has no capacity for conscious volition, and in the view of this author - who grants that there is many an enjoyable philosophical debate to be had on the subject - it is unlikely to acquire any such capacity. But in another sense, there is a message coming through loud and clear from generative AI, if we are willing to listen.

The message is that we need to up our game.

References

[i] Turing, Alan Mathison, 'Computing machinery and intelligence', *Mind*, Volume 59, Issue 236, Oxford University Press, 1950, p433-434

[ii] Ebert, Roger, 'Star Wars - Episode I: The Phantom Menace', *Chicago Sun-Times*, 17 May 1999

[iii] Sohl-Dickstein, J, Weiss, EA, Maheswaranathan, N and Ganguli, S, 'Deep unsupervised learning using nonequilibrium thermodynamics', *Proceedings of Machine Learning Research*, Volume 37, Microtome Publishing, 2015, p2256-2265

[iv] Shakespeare, William, *King Lear* (Quarto 1), 1608, Act 1, Scene 1. Note that in a subsequent edition of the play (the First Folio, 1623) Lear says 'Nothing *will* come of nothing' (emphasis mine).

[v] Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, AN, Kaiser, L and Polosukhin, I, 'Attention is all you need', *Advances in Neural Information Processing Systems*, Volume 30, Curran Associates, p5999-6009

[vi] The relevant formal definition of a function is given in Lejeune Dirichlet, Johann Peter Gustav, 'Über die Darstellung ganz willkürlicher Funktionen durch Sinus- und Cosinusreihen', *Repertorium der Physik*, Volume 1, Moritz Veit and Company, 1837, p152-153. This is collected in Lejeune Dirichlet, Johann Peter Gustav (ed Kronecker, Leopold), *Werke, Band 1*, Georg Reimer, 1889, p135-136.

[vii] The 1837 description is given in Babbage, Charles, 'On the mathematical powers of the calculating engine', collected

in Randell, Brian (ed), *The Origins of Digital Computers*, 3rd edition, Springer Verlag, 1982, p19-54.

[viii] Ada Lovelace's execution trace - headed 'Diagram for the computation by the Engine of the Numbers of Bernoulli' - is an unnumbered foldout page inserted into Note G in her 'Notes by the translator' following Menabrea, Luigi Federico, 'Sketch of the Analytical Engine invented by Charles Babbage Esq', *Scientific Memoirs*, Volume 3, Richard and John E Taylor, 1843, p722-731.

[ix] Undated fragment (thought to have been written before December 1845) of a letter from Ada Lovelace to Lady Byron, collected in Toole, Betty Alexandra (ed), *Ada, the Enchantress of Numbers: A selection from the letters of Lord Byron's daughter and her description of the first computer*, Strawberry Press, 1992, p319

[x] Quotes from Notes A and G in Ada Lovelace's 'Notes by the translator' following Menabrea, Luigi Federico, 'Sketch of the Analytical Engine invented by Charles Babbage Esq', *Scientific Memoirs*, Volume 3, Richard and John E Taylor, 1843, p696, 694, 722

[xi] Turing, Alan Mathison, 'Computing machinery and intelligence', *Mind*, Volume 59, Issue 236, Oxford University Press, 1950, p459

[xii] Markov's original paper on the subject was published in Russian in an edition of the *Bulletin of the Kazan Physico-Mathematical Society* (Series 2, Volume 25) dated 1906 on its title page, but actually published in 1907 (with a 1907 date given at the end of Markov's paper). No English translation is available at the time of writing, but a follow-up 1907 paper in which Markov elaborates on these ideas - 'Extension of the limit theorems of probability theory to a

sum of variables connected in a chain' - is available in English translation as Appendix B to Howard, Ronald Arthur, *Dynamic Probabilistic Systems, Volume 1: Markov Models*, John Wiley and Sons, 1971, p552-576

[xiii] An entertaining account of the dispute is given in Ellenberg, Jordan Stuart, *Shape: The Hidden Geometry of Absolutely Everything*, Penguin, 2022, p84-89

[xiv] Markov Sr, Andrey Andreyevich, 'An example of statistical investigation of the text *Eugene Onegin* concerning the connection of samples in chains', lecture to the Russian Academy of Sciences, 23 January 1913. English translation published in *Science in Context*, Volume 19, Issue 4, Cambridge University Press, 2006, p591-600

[xv] Shannon, Claude Elwood, 'A mathematical theory of communication', *Bell System Technical Journal*, Volume 27, Issue 3, Bell Telephone Laboratories, 1948, p387

[xvi] Boden, Margaret Ann, 'Computer models of creativity', *AI Magazine*, Volume 30, Issue 3, Association for the Advancement of Artificial Intelligence, 2009, p24

LETTERS on LIBERTY

Letters on Liberty publishes regularly. If you want to ensure you don't miss a single one, you can now subscribe and get the next five bundles for just £25 by heading to **www.academyofideas.org.uk/letters**

WHY DEBATING MATTERS

Mo Lovatt argues that debating matters because ideas matter. Rather than shying away from controversial or difficult topics, we should embrace them and encourage young people to do the same.

ABORTION AND THE FREEDOM TO FORGE OUR OWN FATE

Ann Furedi argues that the future of a woman's pregnancy should be for her alone to decide. We cannot respect the principles of freedom without acknowledging the freedom of reproductive choice.

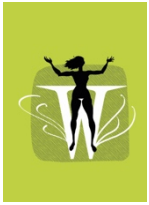
THE TRANS IDEOLOGY TRAP

James Esses argues that instead of broadening the scope of personal liberty, gender ideology actually hinders it, detracting significantly from the rights and freedoms of others - particularly in terms of free speech.

RETHINKING ANTI-SEMITISM

Daniel Ben-Ami argues that a future of freedom from anti-Semitism and the hatred of Jews can only be achieved through a commitment to free and open debate.

THE CASE FOR WOMEN'S FREEDOM



Ella Whelan argues that the fight for women's freedom is being thwarted by a focus on safety and protection. Only through taking on the challenge of living freedom - accepting responsibility for our actions, whether they turn out to become triumphs or mistakes - will women achieve real freedom.

Author

Sandy Starr is deputy director of the Progress Educational Trust (PET), a charity that improves choices for people affected by infertility and genetic conditions. He serves on the oversight group of the project Governance of Stem-Cell-Based Embryo Models, coordinated by Cambridge Reproduction. Previously, he served on the working groups that produced the clinical practice guidance *Ethical Issues in Prenatal Genetic Diagnosis* (2022) and *Prenatal Diagnosis and Preimplantation Genetic Testing for Germline Cancer Susceptibility Gene Variants* (2023). He has written about genome editing in the *British Medical Bulletin*, the *European Journal* and *Microbiology Today*.

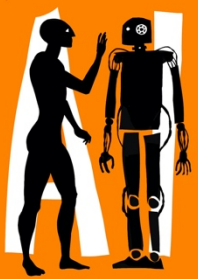
Illustrations
Jan Bowman is an artist and author of *This is Birmingham*. See her work at janbow.com

Letters on Liberty identity
Alex Dale

Pamphlet and website design
Martyn Perks

LETTERS on LIBERTY

academyofideas.org.uk/letters



ISBN 978-1-7395922-9-5