

Between ChatGPT and the singularity: where do we go from here?

Artificial Intelligence (AI) has become increasingly integrated into contemporary society, permeating our everyday lives, entering public consciousness like never before. The advent of generative AI, with the release of ChatGPT3.5 in particular, has opened up the conversation to society more broadly and led to both [greater investment](#) and interest, but also fear and concern. This fear surrounding AI, often fueled by science-fiction and sensationalism, has led to calls for caution, regulation, and proactive protection against potential future dangers. More often than not, [news stories](#) and the public imagination are captured by the idea of Artificial Superintelligence (ASI) and the singularity, where AI overtakes human intelligence, gets out of control and threatens humanity. Comparisons have been drawn between the potential risks of AI and [nuclear war](#). However, unlike nuclear weapons, which are predominantly controlled by state actors, AI can be found in the devices and applications used by everyday individuals. This widespread accessibility poses unique challenges and necessitates careful consideration of how to prevent potential harms. However, while clearly alarming—especially when this future is forewarned by the very [developers of AI themselves](#) and others [call for a pause](#) to any further development—focussing too heavily on this dystopian future will distract us from the present-day harms that affect us already. Moreover, if we address the challenges posed by the current forms of AI today, referred to as Artificial Narrow Intelligence (ANI), it might better prepare us for these potential scenarios in the future.

The risks

ANI presents a variety of challenges that demand our attention, the following points provide a rough landscape of some of the major problem areas.

Discriminatory AI

One is that of biased algorithms and applications of AI that perpetuate racism, sexism and other types of discrimination and oppression. AI systems trained on biased data can [unintentionally reinforce discriminatory practices](#) and amplify social inequalities, since its outputs are based on data points that reflect a world which is itself historically discriminatory and unequal. Secondly, certain applications of AI tend to lead to inequitable outcomes and harm already marginalised groups, for example facial

recognition software that disproportionately [misidentifies black faces](#) — this is particularly problematic when used by [police forces](#) or within justice systems.



Facial recognition software has historically misidentified black faces, leading many to question the usefulness and ethics of such applications.

Misinformation

Another concern is the proliferation of false information online and AI-generated hallucinations that pose risks to the integrity of our information ecosystem.

We are entering an age where it is even more difficult to verify the accuracy of not just text, but also [images, videos and audio](#).

Generative AI makes it possible to create almost any type of content that could be mistaken for real, with potentially devastating effects on news, education, politics and generally life online. Furthermore, AI can inadvertently output false information, so-called '[hallucinations](#)', and the more we rely on AI-generated content, the more we succumb to [automation bias](#) (where we over rely on automated decision making systems and aids), the more at risk we are of being misled and misinformed.



This widely circulated AI-generated picture is of Trump being supposedly arrested as a result of covering up his payment of 'hush money'.

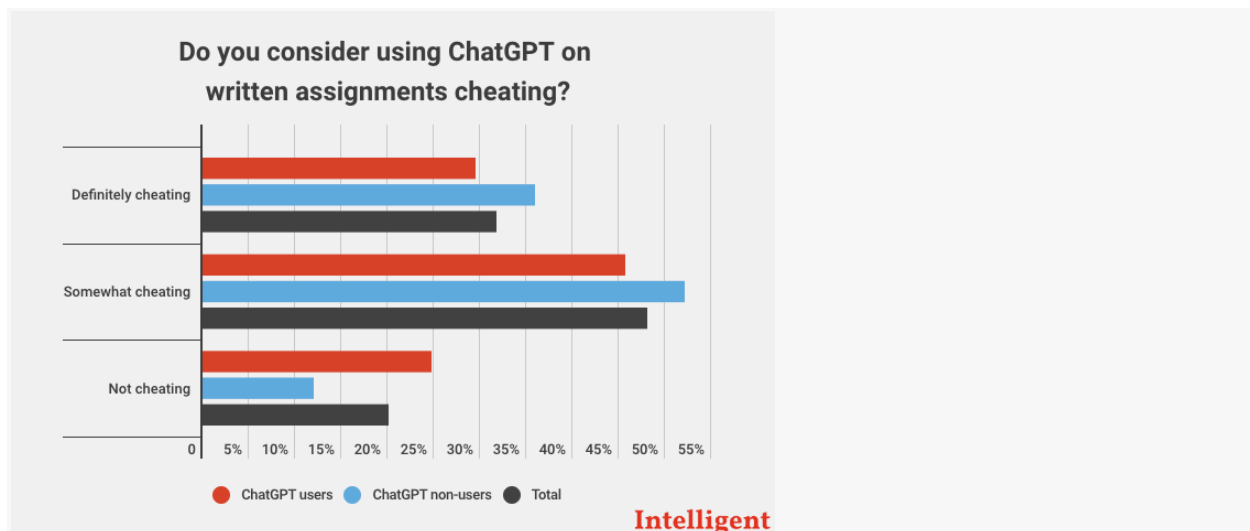
Polarisation

The phenomenon of echo-chambers and rabbit holes is being fueled by AI-driven content recommendation algorithms, which further exacerbates societal divisions and limits the exposure to diverse perspectives. These systems used by platforms such as *Facebook, Twitter, TikTok* and *Youtube* are guided by the desired outcome of more and longer engagement, more clicks, and therefore higher ad revenue. An integral part of [surveillance capitalism](#), this incentive structure necessitates that we see more and more content that will make us engage more—[whether it's something we already agree with or not](#), as long as it drives a strong emotional reaction (and therefore captures your attention for longer). This means that we end up with a warped, [polarised](#) view of society, overestimating (and therefore entrenching) both the popularity of our own views and those that we fundamentally disagree with.

Cheating

No sooner after the launch of ChatGPT3.5 did voices quickly arise [fearing for the state of the education system](#) in its aftermath. While study aids, essay factories and plagiarism are nothing new, the scope and power of AI applied in educational settings

raises concerns about the ease at which students can cheat in their coursework. [One study](#) of 1000 US college students found that 30% had used ChatGPT for their written work, even though three quarters of those users believed it counts as cheating. Since ChatGPT is widely accessible to anyone with a browser and an account, the barriers to using the tool are low, and the temptation for students is high. The impact of this on the quality of their education, and the difficulty for educators to assess (and indeed help improve) the work of their students is potentially profound. Questions remain whether new [software](#) that detects [AI generated content](#) will be sufficient or whether this truly represents a great risk to the next generation of learners. Perhaps it may prove to be merely an evolution of what it means to be a student—that [learning how to use AI](#), as with many new technologies, is integral to education and [institutions should support this](#).



Students overwhelmingly consider using ChatGPT as cheating, regardless if they use it themselves or not.

IP and personal data

The use of artworks as training data has implications for artists and their ability to protect their livelihoods, reputation and their [intellectual property](#). For better known and more prolific artists, this technology is particularly dangerous as there is a greater chance to replicate their style, given the greater quantity of training data, thereby threatening the status and value of their work. There have even been theories gaining traction that the concept of copyright is itself an [aberration](#), and that society will discard it soon. In the meantime, [artists have taken AI companies to court](#) for using their art as training data without permission or attribution.

The Writer's Guild of America also grabbed headlines around the world by [protesting the use of AI tools](#) in script development, which threatens the livelihoods of their members (and, they argue, the quality of film and television writing). However, these ethical considerations do not end with artists and copyrighted material, but all of us by virtue of our data online. This [data is scraped](#) and fed into AI systems without prior consent or knowledge, leading to [allegations of broken privacy laws](#) (when *personal* data is scraped) and bringing up questions about the [potential negative impacts](#) it might have for individuals and businesses. Namely, fears of being defrauded, impersonated or inadvertently enabling hackers to gain access to otherwise secure sites and information.



The Writer's Guild of America took to the streets of Hollywood to protest the use of AI in their industry.

The solutions

As AI makes leaps and bounds, progressing by the day, interested parties, activists, and regulators are playing catch-up, seeking to guardrail the development of AI for the better. Given the involvement of big tech companies, the varied impact across society, and the complex political and geopolitical issues surrounding AI, there can be no single approach in tackling these problems. Indeed, there are several avenues to address them. Individually, they suffer from limitations and unique obstacles, but collectively

they will hopefully take us a long way in grappling with the issues both today and in the future.

Ethical AI research

Firstly, ethical AI research, especially when embedded into product and engineering teams at AI companies. The idea here is that the companies developing AI should invest in ethical AI research teams who are empowered to advise on how AI should be designed and developed. These are individuals trained in data science, moral philosophy and the social sciences, working within companies to steer AI development to better outcomes. However, this is not a task that can be solely entrusted to tech companies, as their profit-driven motives may compromise the integrity and objectivity of this research. The recent rounds of [layoffs](#) in the AI ethics research teams within major tech companies demonstrates this point. This highlights the problem that when the financial outlook worsens, the ethical AI research team is not deemed profitable enough to protect. Infamously, Google was also widely criticised for [firing](#) Timnit Gebru, an influential AI ethics researcher, after she published a paper criticising large language models (which are integral to Google's business model). This incident brings to light the power that tech companies have over their researchers and the fact that they do not enjoy academic freedom, unrestrained from corporate interest or procedure. As a result, independent oversight and robust ethical frameworks are also necessary to complement these teams and ensure that AI development aligns with societal values and principles.

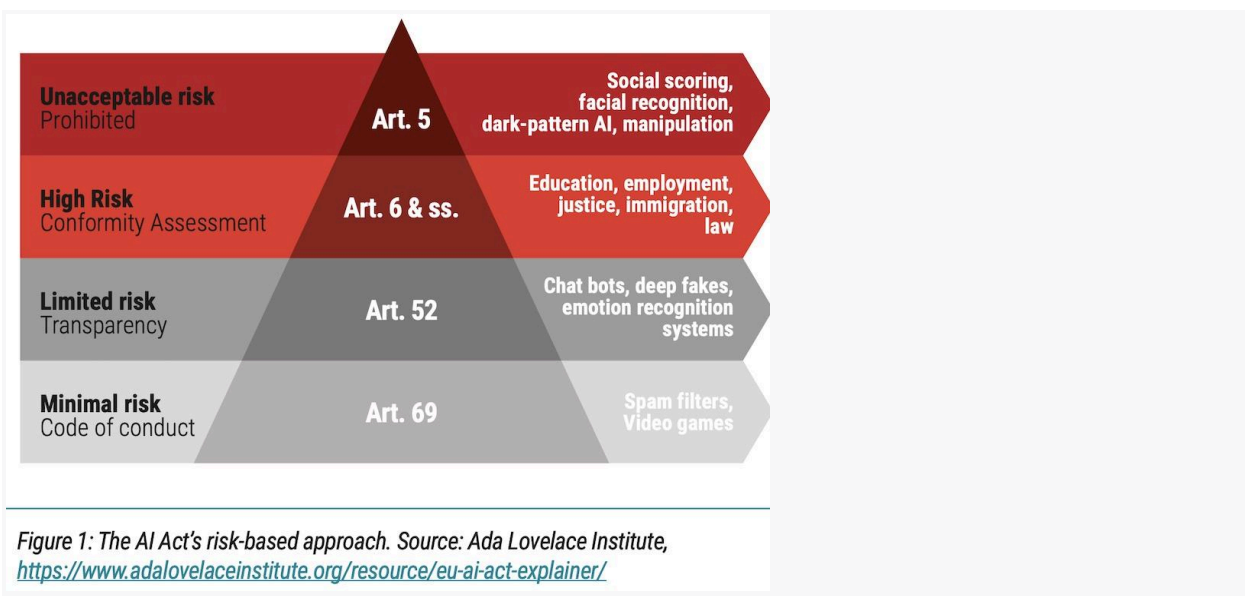


Timnit Gebru's firing brought many to question the true independence of ethical AI researchers working within tech companies.

Regulation

This brings us to regulation, which plays a crucial role in shaping the development and deployment of AI. The European Union recently drafted the AI Act, which aims to

establish comprehensive rules and guidelines for AI technologies. It is a risk-based approach that places stricter requirements (or outright bans) on more problematic applications of AI. For example, using AI as part of a social scoring system (as rolled out in [China](#)) has been banned. It is hoped that this legislation will [set the standard](#) for other countries around the world, in virtue of the [Brussels effect](#), whereby EU regulation impacts the rest of the world. While this is a powerful lever, tech companies still lobby to tailor the regulation to suit their needs. Indeed, the EU was accused of [watering down the regulation](#) as a result of lobbying by *OpenAI*, who argued chatbots should be considered less risky than originally proposed. As a result, regulation such as this tends to [disproportionately favour incumbents](#), who have the funds and political capital to not only influence the regulation but also to cover the cost of compliance. In the United Kingdom, [attempts](#) are also being made to position the country as a leading AI regulator. Though, it would be one with a “[pro-innovation approach](#)” (read: a neoliberal, business-first approach) that arguably will prioritise economic interests over broader societal considerations. We have already seen that the UK [attempted](#) to make an exception to copyright law in order to enable AI companies to use copyrighted data for their training data (before [U-turning](#)). While striking the right balance between regulation and innovation is crucial to harness the potential of AI, it is imperative to do so while minimising its negative consequences: developing AI must not be pursued at any cost. These instances go some way in justifying why we cannot rely on regulation alone. Furthermore, regulations provide a floor, not a ceiling, when it comes to ethical consideration, deliberation and standards. We should strive for more than just avoiding the worst excesses.



The EU AI Act designates various AI applications along a gradient of risk with accompanying requirements.

Education

Education is another essential ingredient in addressing the challenges posed by AI. It is imperative that the debates surrounding AI and its impact is participated in by all levels of society, including engineers, designers, developers, deployers, and everyday individuals, which means including AI in the syllabus. A valuable case study is the approach of Germany in its reckoning with its past (*‘Vergangenheitsbewältigung’*), where successive generations are taught to thoroughly engage with the history of Nazism and strive for a future free from similar atrocities. A common refrain from Germans today is ‘never again’. By [improving AI literacy](#) and instilling a culture of responsible AI development and usage across society over generations, we can aspire to a future where harmful AI practices are prevented from taking root—a new mantra could be adopted: ‘never to begin with’.

We have already seen for a number of years now that computer science departments in the [US](#) and [Europe](#) are running ethics courses for their students. In these, they engage with the moral issues surrounding personal data, AI, and other problematic uses of technology such as surveillance. While this is an encouraging first step, the conversation needs to be broader as it will be not only computer science graduates that design or deploy AI systems in the future.

Inclusion

Lastly, elevating and including the voices and interests of underrepresented and marginalised groups lies at the heart of navigating the intricate landscape of AI. It is crucial to prioritise the voices of those who have historically been marginalised, including women, ethnic minorities, LGBTQ individuals, disabled people, and children. Recent history already shows us how these groups are [disproportionately impacted](#) by AI and are given [fewer opportunities](#) to sit at the tables where relevant, important decisions are made that will affect them. This requires a commitment to inclusivity and the active participation of underrepresented groups. By fostering an environment that encourages their meaningful engagement, we can collectively shape AI technologies that address the unique challenges and concerns faced by these communities. Furthermore, it is crucial to empower and support these groups within civil society, AI companies, and regulatory bodies, ensuring their voices are heard and their interests are protected. The aim is to ultimately align AI systems with the values of society, so that it operates in a way congruent with our ethics, norms, cultures and interests. By

prioritising these voices, we aim to align AI with those who are likely most affected by it.

With that in mind, achieving *global* alignment in the realm of AI values is a formidable challenge. The fragmented nature of the AI ecosystem, with competing systems from around the world attempting to align with different norms and values, further compounds the complexity. Cultural nuances, regional differences, and diverse ethical perspectives influence the development and deployment of AI systems. This reality raises a thought-provoking question:

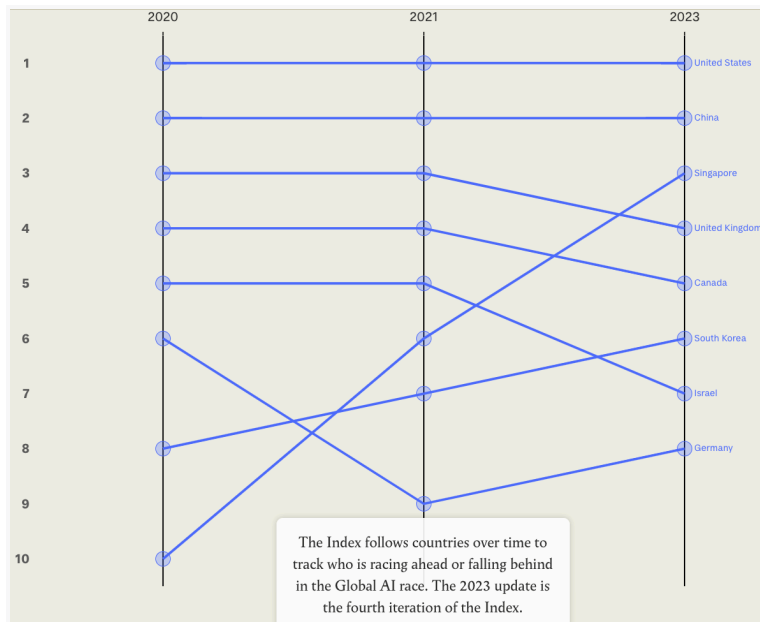
**Can we truly expect to align on a single,
universally accepted set of ethical values in the
context of AI?**

**Is it not more likely that we will end up with
divergent AI's, reflecting the cultures in which
they were developed?**

Perhaps, rather than seeking a single ethical consensus, we can strive for an inclusive and pluralistic approach that acknowledges and navigates the complexities inherent in a fragmented, global AI ecosystem.

The system

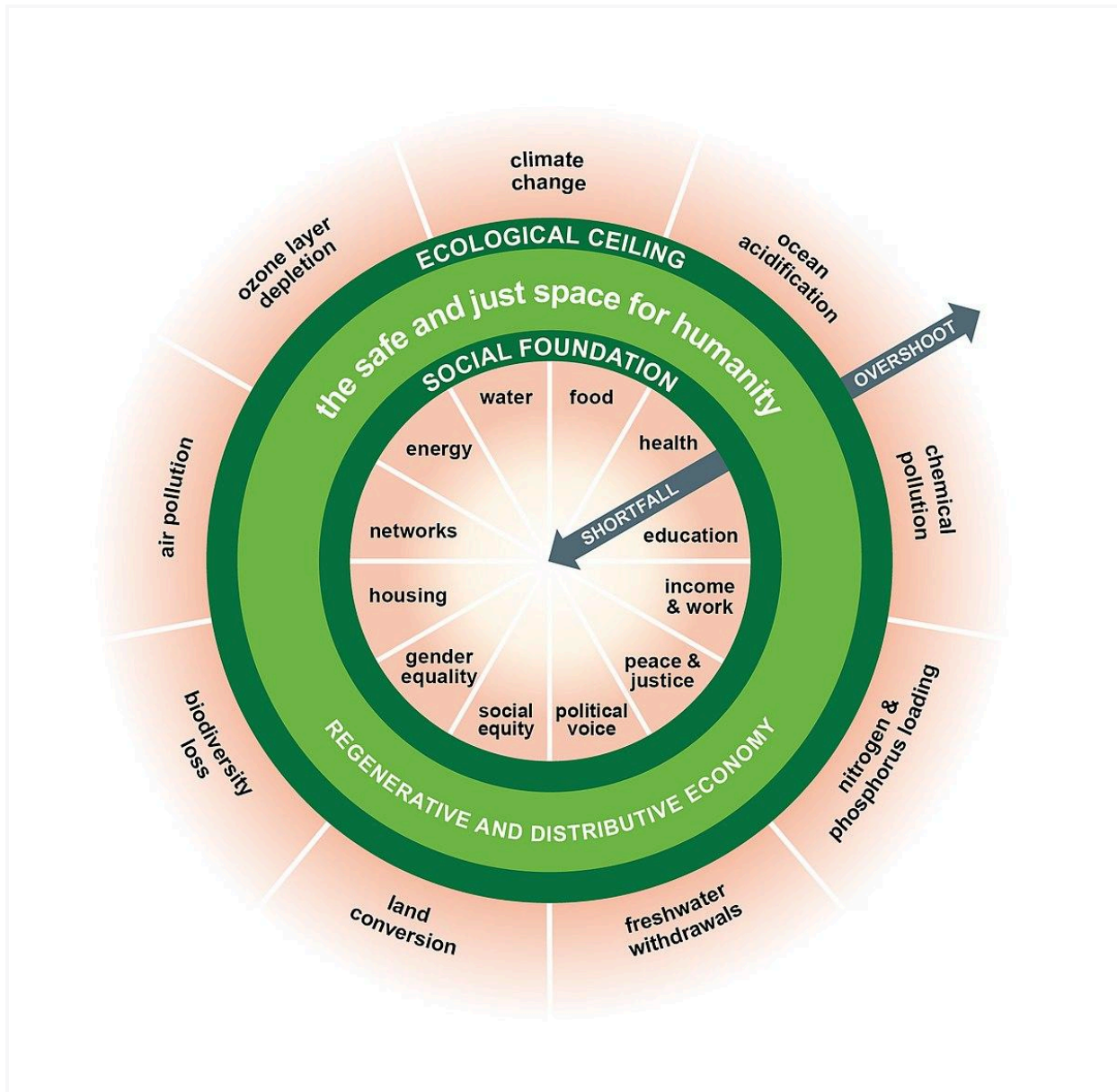
Central to the challenges posed by AI is the fact that it exists within a globalised, predominantly capitalist world. Profit often takes precedence over ethical considerations, rules, and harm reduction. *OpenAI* was itself originally a not-for-profit company, committed to transparency and being open source. In pursuit of further capital, they are now a '[capped profit](#)' company, and a [lot less open](#), keeping ChatGPT datasets and [source code private](#). Even with the grandest ambitions of saving humanity from ASI, the economic realities of living under capitalism will eventually come to bear on those within it. Furthermore, navigating these complex problems becomes particularly challenging when those responsible for building AI systems prioritise competitiveness and profitability. This dilemma mirrors the [coordination problem](#) faced in tackling global warming, where acting in unison could prevent catastrophe, but proves difficult as states seek to preserve economic growth, outweighing the urgency of addressing environmental degradation. This is a problem that cuts across not only the corporate landscape, but on the geopolitical stage too, where regulation can be seen as a [inhibitor to innovation](#) in the context of an [international AI arms race](#).



The global AI arms race has seen the superpowers of USA and China in fierce competition, with great movement amongst smaller players. ([Source](#)).

If we struggle to coordinate efforts to preserve the planet and protect marginalised communities over profit-driven interests today, how can we hope to effectively address the risks posed by ASI in the future?

If companies and countries are racing to be the first to develop the most advanced AI, with profit and domination as the driving factors, then we hasten the inevitability of the ASI worst-case scenario, with few constraints or mitigating measures. It is crucial to create conditions that move beyond capitalism and prioritise the well-being of people and the planet. Initiatives like [B Corp](#), which aim to balance profit with social and environmental considerations, provide us with a starting point. However, it is essential to go beyond superficial [greenwashing](#) and ensure genuine commitment to sustainable and ethical practices. We must be prepared to envision new ways of organising and managing society, stretch our thinking and challenge the status quo. In relation to the ecological crisis, the [doughnut economic model](#) is a useful example of how we might rethink how to live. Similarly with AI, we need big-picture thinking to help us navigate the future if we are to avoid catastrophes on multiple fronts. It is unlikely for there to be a safe, AI-powered future under the systems we have in place today.



The doughnut economic model illustrates the 'safe and just space for humanity', satisfying our social needs within our ecological means.

The promise

You may be forgiven in thinking that the outlook is dim, that AI is fraught with risk and minimal upside. It is important, also, to capture the optimism that many have in the promise of AI and all that it could help us achieve. For there are both tangible and fantastical benefits that AI enthusiasts are keen to celebrate and work towards.

While AI's impact can already be felt across society and various industries, here are a few highlights to give a flavour of the achievements and optimism for the future:

- AI has long been heralded to change the future of work, with some dreaming of a post-work, prosperous, society of leisure (AKA [fully automated luxury communism](#)), and others left fearing for their jobs. [Already](#), we have seen how generative AI is being adopted across industries, automating boring tasks, increasing productivity and serving as a springboard for new ideas. With new technologies come new careers, and, with greater efficiencies, it is hoped that AI will reduce the number of hours we spend on menial tasks at work.
- *DeepMind's* AlphaFold has made [great leaps](#) in the field of Biology, applying AI to the structure of proteins, making accurate predictions and searchable 3D visualisations. The [promise](#) is that this will help us discover new drugs, roll out more vaccines and aid in the fight against antibiotic resistance.
- Despite the concern around the impact on education, there have been [predictions](#) that AI will revolutionise the very methodology of learning by democratising personalised tutoring at a mass scale, around the world. Such a [digital curriculum](#) could profile students as they learn, tweaking their education to match their needs, helping ensure maximal value for students and fostering more of their full potential.
- In the era of climate change, AI holds great promise in slowing down the unfolding catastrophe. Through optimising grid systems, reducing energy inefficiencies, and improving forecasting and predictions, AI is already playing a part in tracking and [reducing emissions](#).
- AI has been rolled out in [cancer screening and imaging](#) processes, helping reduce the time needed to identify cancer, to a high degree of confidence. The appropriate role of AI is an on-going conversation, but at the moment it is verified by radiologists and has been particularly fruitful in breast and lung cancer cases. The hope is that these systems will improve, driving further efficiency and leading to better outcomes for patients over time through 'precision oncology' (the selection of a patient's therapy based on their tumour's molecular profile).

The state of AI in contemporary society demands our attention and proactive measures. While concerns of an AI-induced doomsday scenario capture the public imagination, it is vital not to neglect the pressing challenges posed by current forms of AI. By addressing the biases, misinformation, ethical concerns, and regulatory needs associated with ANI, we can also better prepare ourselves for the potential risks of ASI. This requires ethical research, effective regulation, comprehensive education, and meaningful inclusion of stakeholders. Moreover, as AI operates within a globalised, mostly capitalist world, reevaluating our values and prioritising the well-being of people and the planet are essential steps toward fostering a responsible and sustainable AI-driven future. Only by transcending profit-driven models and embracing holistic approaches can we navigate the complex landscape of AI and its potential impacts on society.

Morgan Williams is a researcher with a background in philosophy and a keen interest in AI ethics. He works for [Stripe Partners](#), a strategy and innovation consultancy based in London, and through their work places people's worlds and cultures at the heart of developing new technologies. He has worked with companies such as *Spotify*, *Meta*, *Ford*, and *1Password*, working with teams in product, design, marketing and research. Engaging in the AI ethics community since 2019, Morgan launched a blog on the subject ([aithics.co](#)) to capture and integrate the technological zeitgeist through a combination of analytic philosophy and ethnographic practice.

morgan.williams@stripepartners.com

<https://www.linkedin.com/in/morgan-l-williams>

<https://www.aithics.co>