Topic

Resolved: The development of Artificial General Intelligence is immoral.

Topic Overview

Artificial General Intelligence - Artificial general intelligence (AGI) is defined as the intelligence of machines that allows them to comprehend, learn, and perform intellectual tasks much like humans. AGI emulates the human mind and behavior to solve any kind of complex problem.
https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-general-ai/

The history of AI development dates back to the 1950s and 60s, but common use AI has really exploded in the last decade or so. A 2024 survey by Pew Research found that 55% of Americans said they regularly use AI.
https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/

Tech companies are spending more and more time and energy on developing AI. Meta, Amazon, Alphabet and Microsoft intend to spend as much as $320 billion combined on AI technologies and datacenter buildouts in 2025.
https://www.cnbc.com/2025/02/08/tech-megacaps-to-spend-more-than-300-billion-in-2025-to-win-in-ai.html

Affirmative

I stand in affirmation of the resolution *Resolved: The development of Artificial General Intelligence is immoral.*

Definitions

immoral: not moral, conflicting with generally or traditionally held moral principles (Merriam Webster)

artificial general intelligence: Artificial general intelligence (AGI) is defined as the intelligence of machines that allows them to comprehend, learn, and perform intellectual tasks much like humans. AGI emulates the human mind and behavior to solve any kind of complex problem. **also called AI

Value-Criterion

fairness: impartial and just treatment or behavior without favoritism or discrimination (Oxford Languages)

This can be evaluated through just representation and access without discrimination.

Con 1: AGI/AI Is Biased

Biased Imagery

Tiku, 2023

Nitasha Tiku, Washington Post, 2023, https://www.washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/

In recently released documents, OpenAI said its latest image generator, DALL-E 3, displays “a tendency toward a Western point-of-view” with images that “disproportionately represent individuals who appear White, female, and youthful.”

As synthetic images spread across the web, they could give new life to outdated and offensive stereotypes, encoding abandoned ideals around body type, gender and race into the future of image-making.

Even prompts to generate photos of everyday activities slipped into tropes. Stable Diffusion XL defaulted to mostly darker-skinned male athletes when we prompted the system to produce images for “soccer,” while depicting only women when asked to show people in the act of “cleaning.” Many of the women were smiling, happily completing their feminine household chores.

Racist Language Models

Miller, 2024

Katharine Miller, Stanford, September 3, 2024, https://hai.stanford.edu/news/covert-racism-ai-how-language-models-are-reinforcing-outdated-stereotypes

Speakers of African American English (AAE) dialect are known to experience discrimination in housing, education, employment, and criminal sentencing. And in a new Nature paper, Kalluri and her colleagues Valentin Hofmann, Dan Jurafsky, and Sharese King demonstrate that covert racism against AAE persists in many of the major large language models (including OpenAI’s GPT2, GPT3.5, and GPT4, Facebook AI’s RoBERTa, and Google AI’s T5). “They generate text with terrible stereotypes from centuries ago, like calling speakers of African American English dirty, stupid, or lazy,” Jurafsky says.

LLM [large language model] developers seem to have ignored or been unaware of their models’ deeply embedded covert racism, Kalluri says. In fact, as LLMs have become less overtly racist, they have become more covertly racist, the Nature paper shows.

As LLMs are incorporated into decision-making systems for employment, academic assessment, and legal accountability, this trend matters. As the researchers showed in additional experiments, compared with users of Standard American English (SAE), LLMs are more likely to give users of AAE lower prestige jobs, more likely to convict them of a crime, and more likely to sentence them to death rather than life for committing a murder. “These results show that using LLMs for making human decisions would cause direct harm to speakers of African American English,” Jurafsky says.

Criminal Sentencing

Noble, 2023

Noble, NPR, July 19, 2023,
https://www.npr.org/2023/07/19/1188739764/how-ai-could-perpetuate-racism-sexism-and-other-biases-in-society

The most life or death dimensions of AI - biased AI, discriminatory AI - would be the use of AI in things like criminal sentencing and, you know, determining whether a person is likely to be a risk or not and keeping them in prison or in jail or releasing them on bail or releasing them entirely.

We saw from the very important research done by Julia Angwin and her team around the COMPAS recidivism prediction software a couple of years ago how Black people who were charged with crimes were more than four times likely to be sentenced to very severe punishment, as opposed to white offenders who were committing violent crimes and were much more likely to be released on bail.

What is used to determine these kinds of predictive AIs are things like histories of arrests in a certain zip code. So if you live in a zip code that has been overpoliced historically, you are going to have overarresting. And we know that the overpolicing and the overarresting happens in Black and Latino communities. That's just a fact. So if that is a main factor in whether you are likely to commit - in predicting whether you're likely to commit another crime because lots of people in the zip code you live in have been arrested more than, let's say, you know, in South Central LA where I live versus in Beverly Hills, then you are more likely to be considered a risk. That has nothing to do with you. That has to do with the history of structural racism in policing in the United States.

Serious Implications

Nadis 2022

Steve Nadis, MIT News, December 16 2022,
https://news.mit.edu/2022/when-subtle-biases-ai-influence-emergency-decisions-1216

A new study by researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and the MIT Jameel Clinic, which was published last month in Communications Medicine, assesses the impact that discriminatory AI models can have, especially for systems that are intended to provide advice in urgent situations. "We found that the manner in which the advice is framed can have significant repercussions," explains the paper's lead author, Hammaad Adam, a PhD student at MIT's Institute for Data Systems and Society.

A key takeaway of the experiment is that participants "were highly influenced by prescriptive recommendations from a biased AI system," the authors wrote.

The MIT team discovered that decision-makers who are themselves unbiased can still be misled by the recommendations provided by biased models. Medical training (or the lack thereof) did not change responses in a discernible way. "Clinicians were influenced by biased models as much as non-experts were," the authors stated.

Con 2: AGI/AI Disrupts Labor Markets

Disruptive Threat of AI

Deming, et. al, 2025

David J. Deming, et. al, National Bureau of Economic Research, January 2025
https://www.nber.org/papers/w33323

General-purpose technologies (GPTs) like steam power and electricity dramatically disrupted the twentieth-century labor market, but the changes took place over decades. We argue that AI could be a GPT on the scale of prior disruptive innovations.

Immediate Decrease in Job Postings

Demirci et. al, Harvard Business Review, November 2024

Ozge Demirci, Jonas Hannane and Xinrong Zhu, Harvard Business Review, November 11, 2024
https://hbr.org/2024/11/research-how-gen-ai-is-already-impacting-the-labor-market

To conduct our study, we analyzed 1,388,711 job posts from a leading global online freelancing platform from July 2021 to July 2023.Online freelancing platforms provide a good setting for examining emerging trends due to the digital, task-oriented, and flexible nature of work on these platforms. We focus our analysis on the introduction of two types of gen AI tools: ChatGPT and image-generating AI. Specifically, we wanted to understand whether the introduction and diffusion of these tools decreased demand for jobs on this platform and, if so, which types of jobs and skills are affected most and by how much.

Using a machine learning algorithm, we first grouped job posts into different categories based on their detailed job descriptions. These categories were then classified into three types: manual-intensive jobs (e.g., data and office management, video services, and audio services), automation-prone jobs (e.g., writing; software, app, and web development; engineering), and image-generating jobs (e.g., graphic design and 3D modeling). We then examined the impact that the introduction of Gen AI tools had on demand across these different types of jobs.

We find that the introduction of ChatGPT and image-generating tools led to nearly immediate decreases in posts for online gig workers across job types, but particularly for automation-prone jobs. After the introduction of ChatGPT, there was a 21% decrease in the weekly number of posts in automation-prone jobs compared to manual-intensive jobs. Writing jobs were affected the most (30.37% decrease), followed by software, app, and web development (20.62%) and engineering (10.42%).

A similar magnitude of decline in demand was observed after the introduction of popular image-generating AI tools (including Midjourney, Stable Diffusion, and DALL-E 2) were introduced. Within a year of introducing image-generating AI tools, demand for graphic design and 3D modeling freelancers decreased by 17.01%.

Additionally, we noticed that over time, there were no signs of demand rebounding, revealing a growing trend of job replacement. We compared this impact against both typical seasonal demand fluctuations on the job platform and the effects that automation had on traditional labor markets. The impact produced by gen AI tools was significantly more substantial. Comparing gen AI's hits to robotic automation for example, researchers found that a 20-percentage-point increase in robot adoption in French manufacturing led to only a 3.2% decline in industry employment.

Troublesome Future Outcomes

Zarifhonarvar, June 2024

Ali Zarifhonarvar, Journal of Electronic Business & Digital Economics, June 6th, 2024
https://www.emerald.com/insight/content/doi/10.1108/jebde-10-2023-0021/full/html#sec003

The effect that ChatGPT and other forms of generative AI will have on the job market in the long term is difficult to predict and is mostly unknown. Nevertheless, we can think of two possible scenarios: One potential scenario is that the introduction of ChatGPT and other forms of generative AI services would result in an increase in the total number of job vacancies as well as an increase in the wage for those positions. This is because these services are predicted to positively influence productivity, which, in turn, is expected to result in economic growth and an increase in labor demand. Consequently, there will be a rise in both employment and wages, resulting in an environment that is more beneficial for employees.

The alternative scenario is that the increased level of automation that generative AI brings about will reduce the demand for human workers. For instance, chatbots may completely automate call centers that human agents previously ran. In this hypothetical situation, a decline in the demand for labor might lead to reduced employment rates as well as a decrease in wages for certain workers. This may be especially troublesome for workers in industries that are substantially touched by automation.

Con 3: AI Disrupts Human Creativity and Morality

Creativity Threatened by AI

De Cremer, et. al, Harvard Business Review, April 2023

David De Cremer, Nicola Morini Bianzino and Ben Falk, Harvard Business Review, April 13, 2023
https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work

In the face of technological change, creativity is often held up as a uniquely human quality, less vulnerable to the forces of technological disruption and critical for the future. Indeed, behavioral researchers even call the skill of creativity a human masterpiece.

Today however, generative AI applications such as ChatGPT and Midjourney are threatening to upend this special status and significantly alter creative work, both

independent and salaried. These new generative AI models learn from huge datasets and user feedback, and can produce new content in the form of text, images, and audio or a combination of those. As such, jobs focused on delivering content — writing, creating images, coding, and other jobs that typically require an intensity of knowledge and information — now seem likely to be uniquely affected by generative AI.

[A] possible scenario is that unfair algorithmic competition and inadequate governance leads to the crowding out of authentic human creativity. Here, human writers, producers, and creators are drowned out by a tsunami of algorithmically generated content, with some talented creators even opting out of the market. If that would happen, then an important question that we need to address is: How will we generate new ideas?

A nascent version of this scenario might already be happening. For example, recent lawsuits against prominent generative AI platforms allege copyright infringement on a massive scale. What makes this issue even more fraught is that intellectual-property laws have not caught up with the technological progress made in the field of AI research. It's quite possible that governments will spend decades fighting over how to balance incentives for technical innovation while retaining incentives for authentic human creation — a route that would be a terrific loss for human creativity.

In this scenario, generative AI significantly changes the incentive structure for creators, and raises risks for businesses and society. If cheaply made generative AI undercuts authentic human content, there's a real risk that innovation will slow down over time as humans make less and less new art and content. Creators are already in intense competition for human attention spans, and this kind of competition — and pressure — will only rise further if there is unlimited content on demand.

AI Can Enable Immoral Behavior

Köbis, June 2021

Nils Köbis, Jean-François Bonnefon & Iyad Rahwan, Nature Human Behavior, June 3rd, 2021
https://www.nature.com/articles/s41562-021-01128-2

AI agents acting as enablers of unethical behaviour (partners or delegates) have many characteristics that may let people reap unethical benefits while feeling good about themselves, a potentially perilous interaction.

Harm Is Considered More Permissible When Committed By AI

Giroux et. al, Journal of Business Ethics, July 2022

Marilyn Giroux, Jungkeun Kim, Jacob C. Lee, Park, Jongwon, Journal of Business Ethics, July 2022
https://www.proquest.com/docview/2685225679?parentSessionId=2Eeq%2FM0P5pdVqAzrpimDD4HQ%2B2qCbUcPJjuXs7e8fts%3D&pq-origsite=primo&accountid=14677&sourcetype=Scholarly%20Journals

Shank et al. (2019) investigated the role of decision-making structures (individual decisionmaking vs. joint decision-making) involving moral violations by AI and human agents. They found that humans who made a mutual decision with AI were faulted less than humans who made the individual decision. Also, people attributed more permission and less fault to AIs (vs. humans) in joint decision-making structures. Gill (2020) also investigated the dilemma between self-sacrifce and other-sacrifce in the context of AI. The author examined how people resolve the moral dilemmas of protecting themselves versus others in the context of autonomous vehicles. This research showed that people tend to consider harming others (i.e., pedestrian) (vs. self) more permissible with autonomous vehicles (vs. self) as the decision agent, which was induced by people's attribution of responsibility to autonomous vehicles. This pattern of harming others persisted for both severe and moderate harm, but it decreased when injuring multiple pedestrians or when the pedestrian was a child.

Negative

I stand in negation of the resolution *Resolved: The development of Artificial General Intelligence is immoral.*

Definitions

immoral: not moral, conflicting with generally or traditionally held moral principles (Merriam Webster)

artificial general intelligence: Artificial general intelligence (AGI) is defined as the intelligence of machines that allows them to comprehend, learn, and perform intellectual tasks much like humans. AGI emulates the human mind and behavior to solve any kind of complex problem. **also called AI

Value-Criterion

progress: advance or develop toward a better, more complete, or more modern state (Oxford Languages)

This can be evaluated through access to AI and understanding of its uses and capabilities.

Con 1: AI Is Not Inherently Immoral

AI Does Not Have Malevolent Motives

Bloom 2023

Paul Bloom, The New Yorker, November 29, 2023
https://www.newyorker.com/science/annals-of-artificial-intelligence/how-moral-can-ai-really-be

The worry isn't that ChatGPT, Bing, or Bard currently have malevolent motives (they do not), or even that they have the self-serving evolutionary goals that we have, such as

survival and reproduction (they do not). Rather, the concern is of unintended consequences. A classic example contemplates an A.I. that has been instructed to create as many paper clips as possible. At first, the machine's goal will align with the very human goal of tidying up loose papers. But then the A.I. might conclude that it can make more paper clips if it kills all humans, so no one can switch off the machine—and our bodies can be turned into paper clips. Computers may lack the common sense to know that a command—maximize the number of paperclips—comes with unspoken rules, such as a prohibition on mass murder. Similarly, as the computer scientist Yoshua Bengio has pointed out, an A.I. tasked with stopping climate change might conclude that the most efficient approach is to decimate the human population.

AI's Morality Is Affected By the Humans Who Use It

McKendrick and Thurai, Harvard Business Review, September 2022

Joe McKendrick and Andy Thurai, Harvard Business Review, September 15, 2022
https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions

A degree of human involvement is called for in all scenarios involving AI-based decisions. Business and technology leaders need to ensure that their AI systems have the necessary checks and balances — along with consistent human oversight — to ensure that AI is ethical and moral.

Machines and data can be adapted and monitored, but the people building and using AI systems need to be educated and aware of the need for more holistic decision-making that incorporates ethics, morality, and fairness.

In our view, AI still has a long way to go in making the ultimate decisions in real-world life situations that require more holistic, subjective reasoning. It still is merely a factual engine that acts based on probabilities and scores, mostly based on historical data, with no context of the implications of the information it is delivering. AI may make the right decisions based on facts, but may lack the empathy that needs to be part of those decisions. We still need humans in the middle to assess the value of insights and decisions to the welfare of humans, businesses and communities. AI can help with providing decision-making points, but humans must still be involved in making that decision – ultimately, it needs to be augmented intelligence instead of pure artificial intelligence.

Con 2: AI Reduces Human Work Time

AI Reduces Work Time

Chen et. al, November 2024

Chen et. al, npj Digital Medicine, November 30, 2024
https://www.nature.com/articles/s41746-024-01328-w#:~:text=As%20AI%20transitions%20from%20serving,more%20substantial%20reduction%20in%20workload.

Our results demonstrated that human-AI collaboration significantly reduced the workload in image-based disease detection without compromising diagnostic performance. Nearly one-fourth of the reading time was saved when aided by concurrent AI. The reading quantity decreased by 44.47% and 61.72% when AI served as the second reader and pre-screening, respectively. Additionally, the sensitivity increased by 12% and specificity remained unchanged, respectively, after the integration of AI.

While AI concurrently aids in reducing reading time, clinicians are still required to review all images. However, when AI replaces the second reader for arbitration in cases of discrepancy with the first reader, it can reduce the image volume by 44.47%. Furthermore, utilizing AI as a pre-screening method to eliminate cases that do not require human review can decrease the interpretation quantity by 61.72%. As AI transitions from serving as a concurrent aid to a second reader and then to a pre-screening tool, it progressively assumes a higher degree of substitution for human involvement, leading to a more substantial reduction in workload.

Examples of AI Increasing Productivity

Baily and Kane, May 2024

Martin Neil Baily and Aidan T. Kane, Brookings, May 2, 2024
https://www.brookings.edu/articles/how-will-ai-affect-productivity/

There is evidence that generative AI can improve the productivity of less-skilled employees within an occupation or organization. Case studies show that genAI increased productivity of call center customer support agents, software developers, and mid-level professionals.

There are also intriguing examples of AI contributing to scientific advancement. A team of physicists at Princeton used AI to control the plasma in their fusion reactor. AI has helped scientists understand how proteins fold, a key step in improving the understanding of biological processes. If AI can contribute to faster scientific advancement, this will add to productivity growth on top of the direct use of AI in businesses.

In the area of health care, AI can read scans and point out problems. Given lab and examination results, it can suggest diagnoses and treatment protocols. It can reduce the paperwork burden by summarizing doctors' notes and preparing insurance claims. Reducing the cost of health care would have a massive benefit to the economy regardless of how well that is captured in the productivity statistics.

AI Frees Up Time

Sidorkin, July 2024

Alexander M. Sidorkin, AI & Society, July 26, 2024
https://link.springer.com/article/10.1007/s00146-024-02019-6

We propose that as AI increasingly takes over both manual and routine cognitive tasks, humans are liberated to focus on uniquely human qualities such as creativity, agency, and the capacity for joy.

The trajectory of human evolution is marked by a paradoxical and yet profound relationship with technology. AI and other advancements are poised to "end us" only in the sense that they will transform us into a different, perhaps happier, species. These technologies promise to dismantle our obsessive preoccupation with productive labor, liberating us from an outdated conception of human essence and ushering in a new one.

Advanced robotics could free countless individuals from repetitive and physically demanding tasks. This liberation extends beyond the factory floor, with the potential to revolutionize agriculture, construction, transportation, and various other sectors, allowing humans to focus on more creative and intellectually stimulating pursuits.

Con 3: AI Must Be Researched to Be Regulated

Need Regulatory Mechanisms

Trotsyuk, September 2024

Artem A Trotsyuk et. al, September 13, 2024
https://pmc.ncbi.nlm.nih.gov/articles/PMC10977140/

In genetics, AI can estimate inherited susceptibility for traits (i.e., polygenic scores) to enhance disease risk prediction and deploy precision therapeutics (3). However, given that these algorithms are typically built on data from people of European ancestry, they can exacerbate health inequities and lead to employment and health insurance discrimination (4). If unaddressed, unintended harms from dual use applications of AI pose significant risks. Now is the time to create regulatory mechanisms, coordinated across multiple domains and agencies, for research and clinical practice involving AI that protect the principles espoused by the OSTP.

The U.S. Must Regulate to Lead

Meltzer 2023

Joshua P. Meltzer, Brookings, May 22, 2023
https://www.brookings.edu/articles/the-us-government-should-regulate-ai/

The US government should regulate AI if it wants to lead on international AI governance.

Whether and how the U.S. regulates ChatGPT—and AI more generally—will help set the tone globally for AI regulation and how to address AI risks without stifling innovation. Sam Altman (CEO of OpenAI, which makes ChatGPT4) for instance emphasized the importance of international cooperation on issues such as AI licensing and auditing. However, more is needed in the U.S. when it comes to AI regulation, which is a precursor to more effective U.S. leadership on international AI governance.

AI Must Be Tested to Be Regulated

Tobin 2023

James Tobin, House of Lords Library, July 18, 2023
https://lordslibrary.parliament.uk/artificial-intelligence-development-risks-and-regulation/

The potential benefits and harms of AI have led to calls for governments to adapt quickly to the changes AI is already delivering and the potentially transformative changes to come. These include calls to pause AI development and for countries including the UK to deliver a step-change in regulation, potentially before the technology passes a point when such regulation can be effective. The chief executive of Google, Sundar Pichai, is one example of a leading technology figure who has warned about the potential harms of AI and called for a suitable regulatory framework.

AI should be tested robustly within established regulatory 'sandboxes'.

Rebuttals

Higher Pay for Jobs in Competition with AI

Demirci et. al, Harvard Business Review, November 2024

Ozge Demirci, Jonas Hannane and Xinrong Zhu, Harvard Business Review, November 11, 2024
https://hbr.org/2024/11/research-how-gen-ai-is-already-impacting-the-labor-market

In addition to job replacement, we also looked at other demand factors like job complexity and how much employers are willing to pay workers. We measured complexity by the number of skills required for each job post and found that it increased by 2.18% for automation-prone jobs compared to manual-intensive ones post-ChatGPT. Employers' willingness to pay for these automation-prone jobs also went up by 5.71%. These results suggest that after ChatGPT's launch, automation-prone jobs have become slightly more complex, requiring a wider range of skills, and that employers are willing to pay more for these jobs.

Despite concerns about job losses, AI also offers opportunities for job augmentation and productivity gains. Our findings show that jobs requiring AI-related skills, such as those involving ChatGPT, are rising. As job roles evolve to integrate AI tools, companies need to focus on building a workforce equipped with a diverse set of skills that adapt to and complement these technological shifts.

AI Allows for More Creative Work

De Cremer, et. al, Harvard Business Review, April 2023

David De Cremer, Nicola Morini Bianzino and Ben Falk, Harvard Business Review, April 13, 2023, https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work

With reduced barriers to entry, we can expect many more people to engage in creative work. Github's Copilot doesn't replace the human writing code, but it does make coding easier for novices, as they can rely on the knowledge embedded within the model and vast reams of data rather than having to learn everything from scratch themselves. If more people learn "prompt engineering" — the skill of asking the machine the right questions — AI will be able to produce very relevant and meaningful content that humans will only need to edit somewhat before they can put it to use. This higher level of efficiency can be facilitated by having people speak instructions to a computer, via advanced voice-to-text algorithms, which will then be interpreted and executed by an AI like ChatGPT.

AI Can End Human Bias

Fitter and Hunt, July 2024

Fawn Fitter and Steven Hunt, SAP, July 24, 2024
https://www.sap.com/swiss/insights/viewpoints/how-ai-can-end-bias.html

Even though AI learns – and maybe because it learns – it can never be considered "set it and forget it" technology. To remain both accurate and relevant, it has to be continuously trained to account for changes in the market, your company's needs, and the data itself.

Today, AI excels at making unconscious bias data obvious, but that isn't the same as eliminating it. It's up to human beings to pay attention to bias and enlist AI to help avoid it. That goes beyond simply implementing AI to insist that it meet benchmarks for positive results; this step is a critical part of a program to use AI responsibly to generate fair outcomes.

Considering Environmental Concerns

Sidorkin, July 2024

Alexander M. Sidorkin, AI & Society, July 26, 2024
https://link.springer.com/article/10.1007/s00146-024-02019-6

It is important to acknowledge the significant energy demands and potential environmental impact associated with the development and deployment of AI systems. The computing power required to train and run these systems can contribute to the stress on an already heating planet. However, it is also worth noting that AI may have the potential to contribute to energy conservation and the production of cleaner energy. AI-powered systems could help optimize energy usage, improve the efficiency of renewable energy sources, and assist in the development of new, sustainable technologies. While the environmental concerns surrounding AI should not be overlooked, a balanced perspective that considers both the challenges and the potential benefits is necessary.