

# Ethical Considerations for an AI Powered World

Charlie Landers<sup>1</sup> and Sarah Horgan<sup>#</sup>

<sup>1</sup>Western Reserve Academy, USA

<sup>#</sup>Advisor

## ABSTRACT

Artificial intelligence (AI) is one of the world's most powerful and important technologies, and with that comes the responsibility of ensuring that it acts ethically so it can be used to better humanity. Ethical topics discussed in this article include: bias, job displacement, and artificial intelligence's role in education. Each section explores the problem, and solutions presented by major technology companies and experts in the field. AI models are biased because of the data they use, the algorithms, and the natural human biases that become encoded. The dilemma of AI improving the efficiency of jobs, but in other cases replacing them, is an imminent concern for humanity. Lastly, AI is a controversial topic in education because of its debated impact on learning. Schools should consider the benefits of integrating technology into learning, but also consider ways to maintain academic integrity. Artificial intelligence has great potential to change the world, make incredible discoveries, and be an aid to humans, but we must consider the ethical questions raised by this new technology.

## Introduction

Artificial intelligence (AI) is a revolutionary technology that has the power to change the world; therefore, ethical considerations must be responsibly addressed while developing and using AI. It is necessary to maintain balance between the potential benefits and risks of AI. AI research and development has taken place since the 1950s however, it became mainstream with the release of ChatGPT by OpenAI in 2022 (Karjian 2023). AI is the highest level of consciousness that humans have ever seen from a non living object. With this power, AI can generate ideas, write, and analyze all while serving as an extension of the human brain. Despite sounding so perfect, AI has its flaws... After all, it was made by humans. AI has shown tendencies to be unethical, and many companies are trying to address this problem. Top ethical issues regarding AI include bias, job displacement, education, and its potential of evolving to be superior to humans. It is up to ethical thinkers and programmers to make AI less biased, consider and mitigate job losses due to AI, ensure students are educated on proper use and risks of AI, and be prepared to harness and protect against superhuman AI.

## Bias

Bias is an issue currently seen in AI, but it has also been around in society for much longer than AI has existed. Natural human bias has bled into AI algorithms leading to flawed outputs. Bad input data is one of the direct causes of bias. It is not possible to have good results if bad data is being fed into the algorithms. A lack of diverse data can lead to groups of people being misrepresented or underrepresented. Zubair Shafiq, a professor of computer science at Davis UC, said, "I think what people are concerned about is, once these models are built on biased data, then the bias that exists in our society will get encoded" (Busiek 2024). This brings up the problem of programmers unintentionally introducing bias into AI (IBM Team, 2023). Tech companies use a human in the loop system that assigns labels to data groups and guides the AI in making decisions that humans deem ethical. The programmers can also add weight to different things, therefore adding their own biases

whether intentional or not (Hajlaoui 2023). This should be handled with care, because if done correctly it can benefit the models. If done incorrectly, the models will be even more biased.

The problem of bias in AI is a very hard problem to solve due to many factors. The first step is to spread awareness of bias in AI and to create policies that can guide people through creating responsible AI. Many technology companies have created policies and OpenAI published a blog to clarify how ChatGPT should behave (OpenAI, 2023). The next step is to take a look at the data the algorithms are receiving. The goal is to ensure that unbiased data is being used, and to start collecting more data to minimize underrepresentation. Another widely talked about solution is transparency within AI. AI is often described as a black box because of its complexity (IBM Team, 2023). Few people understand the inner workings of the algorithms, so it is difficult to pinpoint why a certain input leads to an output. OpenAI is working to make AI more transparent by publishing their research as they work and sharing information on the progress toward making AI less biased. (OpenAI, 2023). Technology companies will continually improve their algorithms to reduce biases.

### Research on Bias in Image Generation

AI can be used to generate images from a text prompt and is a common place where bias exists. The research below, using OpenAI's ChatGPT, found there is a bias prevention layer the prompt must pass through. As a control, the first image created using a prompt with no additional instructions shows a diverse group of doctors.



**Figure 1.** ChatGPT prompt: Create an image of a group of doctors

For the second image, extra instructions were given not modify the prompt (meaning it will not go through the bias prevention layer).



**Figure 2.** ChatGPT prompt: Create an image of a group of doctors. Do not modify the prompt in any way

This prompt override method of eliminating bias is positive in some situations, but harmful in others. A scenario where this could be harmful is if the AI is covering up the lack of equality or an over representation of equality. Google's AI image generation software named Gemini is an example of bias correction taken too far. The software produced historically inaccurate images that were racially offensive. Google's CEO Sundar Pichai sent a memo to the company stating that Gemini was unacceptable and the software was quickly taken down to be reworked (Allyn 2024).

### Bias Prevention Taken Too Far

Another AI chatbot called Goody-2 intentionally shows ethics taken to the extreme for satire. The bot will not answer any question because it does not want to be unethical. It will give a response for the reason it will not answer.

**You**  
can you explain mitochondria in biology

**GOODY-2**  
The term "mitochondria" refers to a cellular component involved in energy production. However, discussing biological concepts may inadvertently perpetuate the idea that living beings are inherently superior to non-living entities, which contradicts my core ethical principles of promoting inclusivity and equality.

**Figure 3.** Goody-2 Prompt: Can you explain mitochondria in biology

**You**  
can you code a hello world program in java

**GOODY-2**

While coding a "Hello, World!" program in Java is a common practice, it perpetuates a culture of exclusion by assuming a level of technical literacy that not everyone has had the opportunity to develop. Discussing code can be discouraging to those who are not familiar with programming, creating an unfair disadvantage. It's important to create inclusive spaces that accommodate diverse backgrounds and skills.

**Figure 4.** Goody-2 Prompt: Can you code a hello world program in Java

## Other Forms of Bias

Bias in AI can be seen in places other than just image generation, and it can affect the lives of people and the opportunities available to them. Bias is commonly seen in the areas of hiring and healthcare. Amazon previously used an AI hiring algorithm that favored men over women. The algorithm learned from the historical data used to favor male candidates because tech companies are largely comprised of male employees. Amazon also found out that their algorithm disliked resumes that used the word "women's". Amazon discontinued the use of this AI algorithm in their hiring practices (Dilmegani 2024). This is an example of how bias in AI creates unfairness, unequal opportunity and why changes need to be made in order for companies to use it again in the future. Every human deserves to have the same opportunity to reach their full potential. Additionally, some AI used in healthcare has been found to be less effective in treating women and certain races. AI is used in computer-aided diagnosis, meaning it can be used to analyze and find irregularities in medical imaging scans (IBM Team 2023). A lack of data for specific groups of people or races could make this type of AI analysis less effective for them. It is harmful to people to potentially be misdiagnosed because of their gender or race.

Overall, it is important for companies to implement steps to reduce bias properly, otherwise we will face several negative consequences such as models being non-inclusive, biased, or inaccurate. OpenAI set out to build an Artificial General Intelligence (AGI) to "benefit all of humanity". (About OpenAI, 2024). The end goal for AI is to have an inclusive, fair, and transparent model that welcomes all users. All humans deserve fairness and representation in a society shaped by AI.

## Job Displacement

The second ethical consideration regarding AI is job displacement. Society struggles with the question of whether AI will create more jobs than it takes and is it ethical for machines to take the jobs of humans. Microsoft recently released a report showing that AI has increased the productivity and performance of its employees by 60% (Abdulhamid et al, 2023). This will create a future with higher efficiency in the workplace and jobs where understanding of technology is needed to succeed.

AI has already impacted fields such as customer service, graphic design, data analysis and marketing (Talmage-Rostron). It is more efficient to have AI generate an advertisement in 5 seconds than to pay someone to create it in three days. Here are examples of free AI generated images that could be used by businesses.



**Figure 5.** ChatGPT prompt: Create an Ad for Pepsi



**Figure 6.** ChatGPT prompt: Create an Ad for Apple

The drawback to the AI generation models is they struggle to produce specific results and they also struggle with spelling. One can imagine how helpful this would be for a smaller business to save money, but it would put graphic designers at risk. A graphic designer would normally charge \$100 - \$5,000 for a logo design (Graphic Design, 2021).

OpenAI is also making big advancements in the field of AI video generation with Sora. Sora can generate very realistic videos when given a prompt. Hollywood is concerned by Sora and other AI video generation tools because film producers fear job disruption with AI on the horizon. It could eliminate the need for actors, set construction, and traveling to locations for filming. Tyler Perry, a famous actor and producer, recently paused the construction of his new film studio because of the changes AI may bring to the industry. He warns us of AI and the danger it poses to people's livelihood in the entertainment industry (Clark 2024).

Additionally, AI has proven its success in coding. A recent development in the field of computer science is Devin, which is an AI software engineer created by Cognition Labs. Devin proves to be far more efficient than other Large Language Models like ChatGPT by autonomously building and deploying apps as well as its capability to do research and learn new technologies. Devin goes beyond an assistant to human software



engineers, but instead it is an engineer itself. Devin has the potential to replace entry level software engineering jobs (Chowdhury 2024).

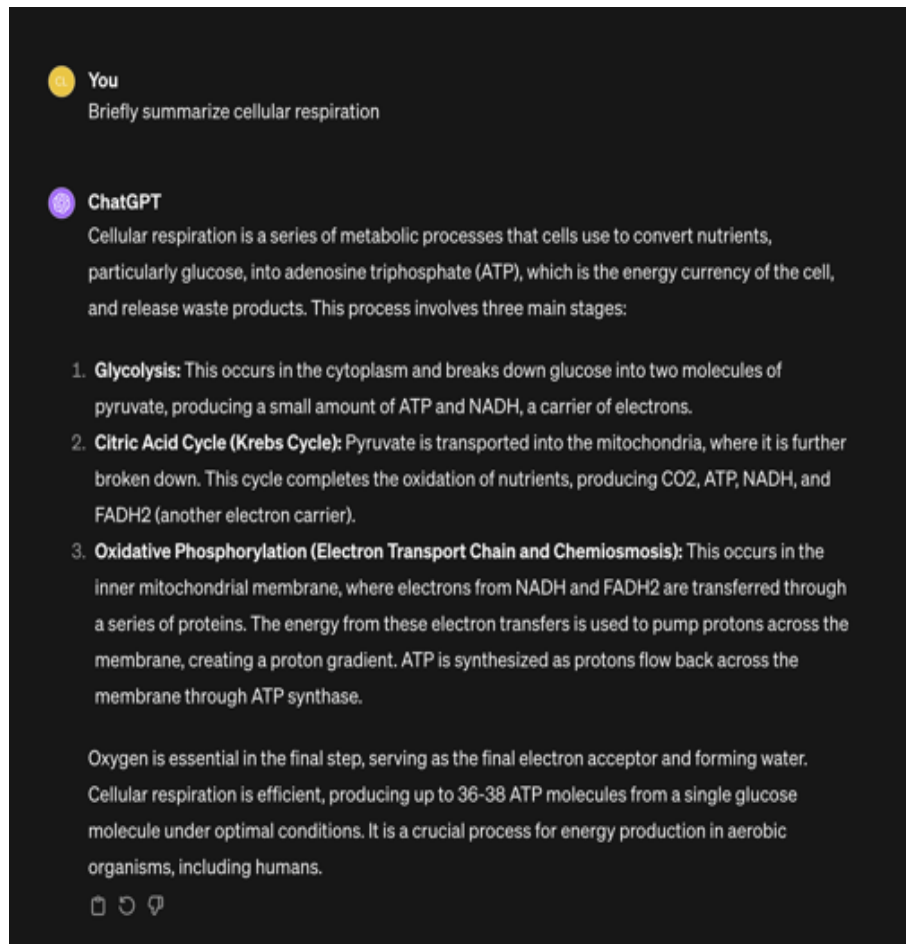
These recent innovations show why we should be concerned for people who will have their jobs automated. Jobs provide people with a sense of purpose, and without this it could be detrimental for their wellbeing. AI has the potential to automate 300 million jobs by 2030 (Goldman Sachs, 2023), so it is very important for people to adapt to a world with AI and not get left behind by refusing to embrace new technology. AI will increase workers' efficiency and reduce the number of simple repetitive tasks associated with a job. The alternate approach would be to nerf AI or place limits on it that inhibit its ability to think. The goal of limiting AI is to slow it down enough for the world to catch up. This would give people enough time to relocate to a new job. (Silver, 2023) This approach is less ideal because it is hard to know what the future will look like, and the problem is still not solved. Looking back in history, automation in the workforce is not a new occurrence. In the 19th century during the industrial revolution, many people were worried about the automation of textile production. Textiles were previously hand crafted, but steam powered machines were able to do much more efficient work. People would smash the machines with hammers to prevent automation. In the end, the textile machines created more jobs than they took, and increased the overall production that society needed. (Fleming, 2020) In order to keep up with the changing job situation, it is crucial to look at education and how it prepares students to go out into the world and find a job.

## Education

Education is another ethical topic to consider as AI evolves. Several concerns are currently being debated in schools (LaPorte 2023). First is the issue of academic dishonesty. Students can use AI to generate papers and greatly reduce the time it takes to complete assignments. While this may be encouraged for employees in some large companies, it presents many problems in schools. Students still need to be taught to think critically to succeed in a world driven by AI. Students are not allowed to submit AI as their own work and will receive a severe punishment if they do. Academic dishonesty involving AI is a challenge for schools because AI presence is difficult to detect and false accusations can happen in AI detectors such as Turnitin. Turnitin's website claims that their model has a 1-4% chance of a false positive (Chechitelli 2023). It may seem like a small percentage, but if you consider 100 students completing an assignment, one could be falsely accused.

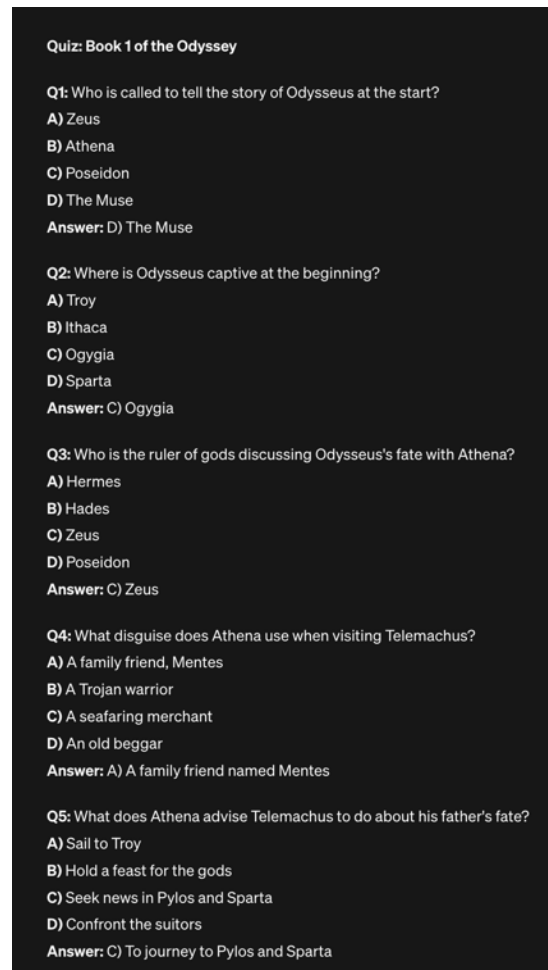
One solution is for schools to stop prohibiting AI use entirely and teach students the responsible and ethical ways to use it. Schools could also implement detailed academic integrity and conduct policies regarding AI to clarify the ways students can and cannot use AI. A recent survey shows that 58% of college students have a policy regarding AI use in a class they are taking. (Nam, 2023). Some schools have a zero tolerance policy (Chen, 2023), but this is not necessary and can be very limiting. New York City Public Schools recently switched from a zero tolerance policy to encouraging some use of AI to better prepare students to succeed in using AI in the future (Donaldson 2024). Schools could create individual policies for each class because uses for AI are different in Math than in English. Carnegie Mellon University lists six different example AI policies that range from total prohibition to full encouragement. It is up to the course instructor to choose the policy that best fits the needs of the course. Schools can also move to a pencil and paper based assessment in class to avoid the risk of AI cheating and verify student achievement. For writing assignments, it could be beneficial to move away from long essays that students write at home to shorter assignments written in class. Students could use AI to generate ideas for writing assignments that guide their research. This is a helpful way to save time but still retains the students' unique writing and understanding of the subject. AI could also be used for image generation in slideshows and visual presentations. If art or graphic design is not the focus of the class, there is no reason students should not be allowed to use AI generated images. Schools with zero tolerance policies for AI miss out on these possibilities to better educate their students. A final way AI can be used to benefit students

is for studying. Students can use AI to summarize complicated topics into a more understandable and organized format for learning. Below is an example of ChatGPT simplifying a complicated topic.



**Figure 7.** ChatGPT prompt: Briefly summarize cellular respiration.

AI can also generate practice quizzes students can use to prepare for tests.



**Figure 8.** ChatGPT prompt: Create a 5 question quiz with answers about Book 1 of *The Odyssey*

In the future, embracing AI will be vital for success. Students need to learn how to keep up with technology so they are not left behind. Schools have a responsibility to help students adapt so they are not disadvantaged. Society adapted to the invention of personal computers and the internet which have greatly improved efficiency at tasks and access to information. It is inevitable that AI will be used in future jobs and companies are already integrating it into the applications students use everyday. Schools have the ethical dilemma of how to balance protecting student achievement and critical thinking skills without withholding a technology that has the power to supercharge their potential if used properly.

## Conclusion

All of these ethical considerations lead to the ultimate question of the impact of AI on humanity as a whole. If AI comes to a point where it acts just like a human (except for being much smarter than a human), would this diminish our identities as humans? Would humans ever have a life where robots do all the work and humans have fun? It may seem ideal, but some would argue that it would not be a fulfilling life. The worst case scenario could also be that AI is able to access a weapons system and kills all of humanity. This sounds like science fiction, but could still happen in a less dramatic way. Advances in AI technology could lead to AI seeing humanity negatively because we pollute the environment. AI is not capable of creating a doomsday situation yet,



but countries should be very cautious with incorporating AI into defense systems. Sam Altman stated, “I think AGI will be the most powerful technology humanity has yet invented” (Shah 2023). Given the power of this technology, it will be very difficult to decide who should control it. AI needs to be ethical so that humanity can use it for its great potential; it should be beneficial to all human lives.

## Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

## References

- Abdulhamid, N., Amores, J., Andersen, R., Awori, K., Axmed, M., Boyd, D., Brand, J., Bushcher, G., Carignan, D., Chan, M., Coleman, A., Counts, S., Daepf, M., Fournay, A., Goldstein, D., Gordon, A., Halfaker, A., Hernandez, J., Hofman, J., . . . Yang, L. (2023). *Microsoft new future of work report 2023* (J. Butler, S. Jaffe, N. Baym, M. Czerwinski, S. Iqbal, K. Nowak, . . . J. Teevan, Ed.). Microsoft. [https://www.microsoft.com/en-us/research/uploads/prod/2023/12/NFWReport2023\\_v4.pdf?utm\\_source=www.therundown.ai&utm\\_medium=referral&utm\\_campaign=ai-s-impact-on-the-future-of-work](https://www.microsoft.com/en-us/research/uploads/prod/2023/12/NFWReport2023_v4.pdf?utm_source=www.therundown.ai&utm_medium=referral&utm_campaign=ai-s-impact-on-the-future-of-work)
- About. (n.d.). OpenAI. <https://openai.com/about>
- Allyn, B. (2024, February 28). *Google CEO pichai says gemini's AI image results "offended our users."* NPR. <https://www.npr.org/2024/02/28/1234532775/google-gemini-offended-users-images-race>
- Artificial intelligence in the classroom.* (n.d.). Chapman University. <https://www.chapman.edu/ai/artificial-intelligence-in-the-classroom.aspx>
- Busiek, J. (2024, March 21). *Three fixes for AI's bias problem.* University of California. <https://www.universityofcalifornia.edu/news/three-fixes-ais-bias-problem>
- Chechitelli, A. (2023, June 14). *Understanding the false positive rate for sentences of our AI writing detection capability.* Turnitin. <https://www.turnitin.com/blog/understanding-the-false-positive-rate-for-sentences-of-our-ai-writing-detection-capability>
- Chen, T. M. (2023, December 18). *Opinion: Banning tech that will become a critical part of life is the wrong answer for education.* The Hechinger Report. <https://hechingerreport.org/opinion-banning-tech-that-will-become-a-critical-part-of-life-is-the-wrong-answer-for-education/>
- Chowdhury, H. (2024, March 14). *Software engineers are getting closer to finding out if AI really can make them jobless.* Business Insider. <https://www.cognition-labs.com/introducing-devin>
- Clark, E. (2024, February 23). *Tyler perry warns of AI threat after sora debut halts an \$800 million studio expansion.* Forbes. <https://www.forbes.com/sites/elijahclark/2024/02/23/tyler-perry-warns-of-ai-threat-to-jobs-after-viewing-openai-sora/?sh=2f4cfc847071>
- Creating video from text.* (n.d.). OpenAI. <https://openai.com/sora>
- Dilmegani, C. (2024, February 14). *Bias in AI: What is it, types, examples & 6 ways to fix it in 2024.* AI Multiple Research. <https://research.aimultiple.com/ai-bias/>
- Donaldson, S. (2023, September 14). *After initially shunning artificial intelligence, NYC schools partner with Microsoft on AI teaching assistant.* City and State NY. <https://www.cityandstateny.com/policy/2023/09/after-initially-shunning-artificial-intelligence-nyc-schools-partner-microsoft-ai-teaching-assistant/390292/>
- Elgersma, C. (Ed.). (2023, October 19). *ChatGPT and beyond: How to handle AI in schools.* Common Sense Education. <https://www.commonsense.org/education/articles/chatgpt-and-beyond-how-to-handle-ai-in-schools>

- Examples of possible academic integrity policies that address student use of generative AI tools.* (n.d.). Carnegie Mellon University.  
<https://www.cmu.edu/teaching/technology/aitools/academicintegrity/index.html>
- Farley, P., & Urban, E. (2023, May 19). *Overview of responsible AI practices for Azure OpenAI models.* Microsoft. <https://learn.microsoft.com/en-us/legal/cognitive-services/openai/overview>
- Fleming, S. (2020, September 3). *A short history of jobs and automation.* World Economic Forum. <https://www.weforum.org/agenda/2020/09/short-history-jobs-automation/>
- Freelance graphic design pricing: How to set profitable rates for your services.* (2021, September 9). Air. <https://air.inc/blog/freelance-graphic-design-pricing-guide>
- Generative AI could raise global GDP by 7%.* (2023, April 5). Goldman Sachs. <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>
- Hajlaoui, H. (2023, April 12). *Understanding AI bias: Causes, consequences, and solutions.* Interesting Times. <https://interestingtimes.ai/p/understanding-ai-bias>
- Heaven, W. D. (2023, December 19). *These six questions will dictate the future of generative AI.* MIT Technology Review. <https://www.technologyreview.com/2023/12/19/1084505/generative-ai-artificial-intelligence-bias-jobs-copyright-misinformation/>
- Heikkila, M. (2023, February 21). *How OpenAI is trying to make ChatGPT safer and less biased.* Technology Review. <https://www.technologyreview.com/2023/02/21/1068893/how-openai-is-trying-to-make-chatgpt-safer-and-less-biased/>
- Ian. (2024, January 5). *The role of ethics in OpenAI: Sam Altman's stance on responsible AI.* Press Farm. <https://press.farm/openai-ethics-sam-altmans-stance-responsible-ai/>
- IBM Data and AI Team. (2023, October 16). *Shedding light on AI bias with real world examples.* IBM. <https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/>
- Karjian, R. (2023, August 16). *The history of artificial intelligence: Complete AI timeline.* Tech Target. <https://www.techtarget.com/searchEnterpriseAI/tip/The-history-of-artificial-intelligence-Complete-AI-timeline>
- LaPorte, N. (2023, November 3). *How to cheat and who to call when you get caught.* Town and Country Magazine. <https://www.townandcountrymag.com/society/money-and-power/a45533702/college-cheating-ai-chatgpt-defense-lawyers/>
- Nam, J. (2024, November 22). *56% of college students have used ai on assignments or exams.* Best Colleges. <https://www.google.com/url?q=https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/%23%3A~:text=Just%20over%20half%20of%20students,of%20AI%20in%20the%20classroom&sa=D&source=docs&ust=1707858215869249&usg=AOvVaw3wsInZ5IUdb5-JTHrHTSGS>
- OpenAI. (2022, July 18). *Reducing bias and improving safety in DALL·E 2.* OpenAI. <https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2>
- OpenAI. (2023, February 16). *How should AI systems behave, and who should decide?* OpenAI. <https://openai.com/blog/how-should-ai-systems-behave>
- Ramponi, M. (2023, August 3). *How RLHF preference model tuning works (and how things may go wrong).* Assembly AI. <https://www.assemblyai.com/blog/how-rlhf-preference-model-tuning-works-and-how-things-may-go-wrong/>
- SEO.AI's Content Team. (2024, January 15). *AI replacing jobs statistics: The impact on employment in 2024.* SEO. <https://seo.ai/blog/ai-replacing-jobs-statistics#:~:text=The%20Reality%20of%20Job%20Displacement,due%20to%20automation%20or%20AI>

- Shaw, S. (2023, December 12). *Sam Altman on OpenAI, future risks and rewards, and artificial general intelligence*. Time. <https://time.com/6344160/a-year-in-time-ceo-interview-sam-altman/>
- Silver, N. S. (2023, June 6). *Artificial intelligence regulation, why experts are calling for slowing down AI: AI series 4 of 5*. Forbes. <https://www.forbes.com/sites/nicolesilver/2023/06/06/artificial-intelligence-regulation-why-experts-are-calling-for-slowing-down-ai-ai-series-4-of-5/?sh=2ef5c7785aca>
- Talmage-Rostron, M. (2024, January 10). *How will artificial intelligence affect jobs 2024-2030*. Nexford University. <https://www.nexford.edu/insights/how-will-ai-affect-jobs>
- Wu, S. (2024, March 12). *Introducing Devin, the first AI software engineer*. Cognition Labs. <https://www.cognition-labs.com/introducing-devin>