Ethics in the Age of Algorithms: Exploring Moral Implications in AI and Machine Learning

Niuni Amarasinghe

Table of Contents

1: Introduction

Artificial Intelligence (AI) and Machine Learning (ML) has had a direct influence in our lives. The advantages inform us how it can revolutionise healthcare, enhance production efficiency, and contribute to education. But the associated risks can be ignored, including the lack of transparency, the apprehension of the unknown, and the negative impact on lives.

In the AI revolution, to understand what it can do and establish clear guidelines of control is imperative. This essay will explore what ethical considerations should be integrated into the design and development process of AI and ML systems and how it can be implemented through literature study. The social, economic, and environmental impact through the influence of employment and economic growth will be investigated using modelling in Mathematica.

There are two research questions:

1. Part A - Why should ethical considerations be integrated into the design and development process of AI and ML systems?
2. Part B - How do AI and ML influence employment and economic growth?

2: Methodology

2.1: Literature Study

Articles were sourced from reputable websites such as Elsevier, and were checked for reputable publishers and professional documentation. The sources needed to be relevant to the research questions, whether through case studies or arguments by using recent publications.

To ensure heterogeneity, it was important to find sources with contrasting perspectives to prevent bias by using peer-reviewed publication which enhances their reliability. Ethical studies require an objective point of view so consulting a range of ideologies in literature is beneficial. Figure 1 shows an example of a simple random network and represents the interconnected network of citations.

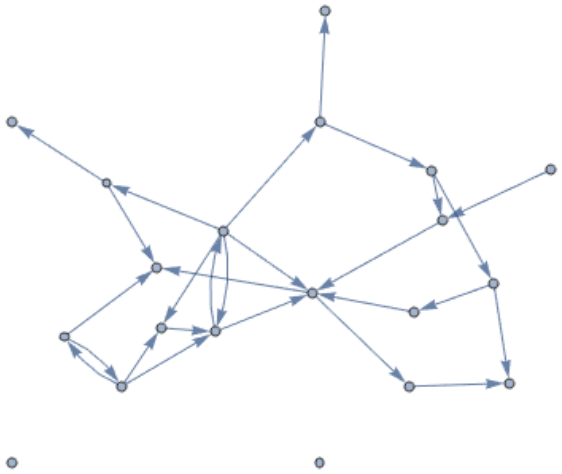

Figure 1: Example of a simple random network[1].

When tracking a network of citations, there is a potential issue of citations being used in the wrong context and being misinterpreted through diffusion (della Briotta Parolo et al., 2020). According to Radicchi, Fortunato and Castellano (2008), field variation in the context of citation analysis is important to consider since there can be a higher number of publications in one discipline compared to another. This means taking the number of citations as an indicator for credibility is ineffective, especially since the field of AI and ML is a fairly new concept in this century compared to, for example, medicine. To overcome this, checking the reputation of the author and publisher was focussed on to ensure that any evidence being cited was valid. There is still a chance of misrepresentation, so an external opinion, for example, ChatGPT (OpenAI, 2023a), was beneficial.

[1] Lecture 3 ECON0055, Dr Frank Witte, UCL, London, 17th October 2023

2.2: Modelling

The first model was a gradient-learning model in the context of AI to understand the impact on employment. The aim was to determine how AI can be used to increase efficiency in a job search, thus changing the market for employment.

The second model attempted to understand how AI would impact economic growth in a country. It was based on the context of Schumpeterian growth since AI effects innovation and creative destruction (Aghion, Akcigit and Howitt, 2015).

With more time, the models in this essay could have included cross-validation, such as by Yoon (2020). Yoon combines a regression tree to the gradient boosting model, enabling new trees to be iteratively made whilst incorporating errors of previous trees until no further improvement is possible. The models in this essay only use the first iteration and so require fine-tuning to be more accurate, such as by using Yoon's (2020) technique.

Coding was implemented by referring to examples completed in lectures and practicals. Amendments were made to make them in the context of AI. By using Mathematica, an interactive system, there was the option to evaluate each code cell as you go, which made it easier to identify when mistakes were being made and also check that the right result was there. If there were any problems, referring to the Wolfram Documentation was helpful to understand the purpose of any functions. If confused, ChatGPT (OpenAI, 2023a) was used, an AI tool, to help diagnose the issue.

3: Results

3.1: Part A - Ethical considerations

A UCL study by Caldwell et al. (2020), identified the crimes that can occur from AI and how they were ranked in the study:

Figure 2: Mind map showing the potential crimes from AI and ML (Cadwell et al., 2020).

3.2: Part B - Influence on employment and economic growth

3.2.1: Gradient-learning model of using AI (employment)

Scenario: 'You are applying for a job in the current economic market, and are using AI to increase your efficiency of submitting applications'.

$$\frac{dJ}{dt} = \alpha U[J[t]] - \beta J[t]^2$$

α a positive constant indicating the effectiveness of using AI to improve the application and β is a positive constant indicating diminishing returns on job application efficiency over time, e.g. level of efficiency you can maintain without 'burning out'.

The utility function is defined as:

$$U[J[t]] := mJ[t] - \frac{c}{2} J[t]^2$$

[2] Lecture 6 ECON0055, Dr Frank Witte, UCL, London, 14th November 2023

$J[t]$ represents the number of job applications submitted over time, m is a positive constant represent the monetary gain or satisfaction from applying, and c is a positive constant representing the cost of time and effort spent on resources and submitting applications. The quadratic cost function indicates that costs will increase as the number of applications increases.

Finding the solution to the differential equation for a gradient-learning person using AI to aid in their job search and solving with initial conditions enables plots as in Figure 3. To maintain consistency, m and c were equal in every scenario.

	α	β	**Meaning**
1	Large	Large	AI has a significant impact on improving application efficiency, represented by a large increase in utility but has rapid diminishing returns on job application efficiency over time since efficiency decreases quickly over time.
2	Small	Small	AI's impact was lesser on application efficiency and there is slower diminishing returns on job application efficiency over time.
3	Large	Small	AI has a significant impact on improving application efficiency and there is slower diminishing returns on job application efficiency over time.
4	Small	Large	AI's impact was lesser on application efficiency but has rapidly diminishing returns on job application efficiency over time.

Table 1: The four scenarios depending on positive constants α and β.

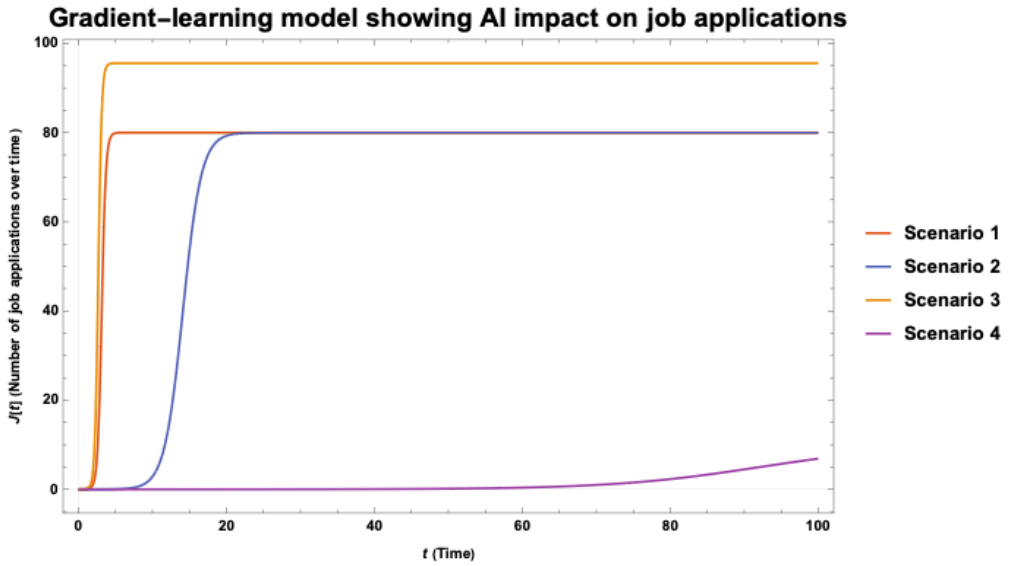

Figure 3: Plot showing the gradient-learning model and the scenarios.

Assuming that when job applications reach their peak, a job is secured, Scenario 1 leads to the best outcome, with the least amount of time spent and number of applications. Scenario 2 is similar, but takes more time to reach the same outcome. The next best is Scenario 3, with about 10 more applications needing to be submitted in slightly less time. Scenario 4 is the worst-case scenario, where AI is not being efficiently used by the user.

3.2.2: AI and Schumpeterian growth (economic growth)

This model simulates the implementation of AI leading economic agents to apply it throughout the whole economy via gradient-learning. The utility function indicates the knowledge and efficiency gained from AI, x:

$$U[x] = \alpha\,(A\,x + B\,x^2)^3$$

α is the utility scaling factor, A is the coefficient for knowledge gained from AI and B is the coefficient for efficiency gained from AI.

The gradient-learning equation would be:

$$\partial_t x[t] = L\partial_{x[t]} U[x[t]]$$

$x[t]$ is the variable representing the knowledge and efficiency gained from AI over time and L represents the learning rate of the economy.

Over time, the use of AI can increase the level of innovation, creating a time-series of knowledge equilibria. This model simulates AI and ML's impact on total factor productivity (TFP) in a developed and developing country.

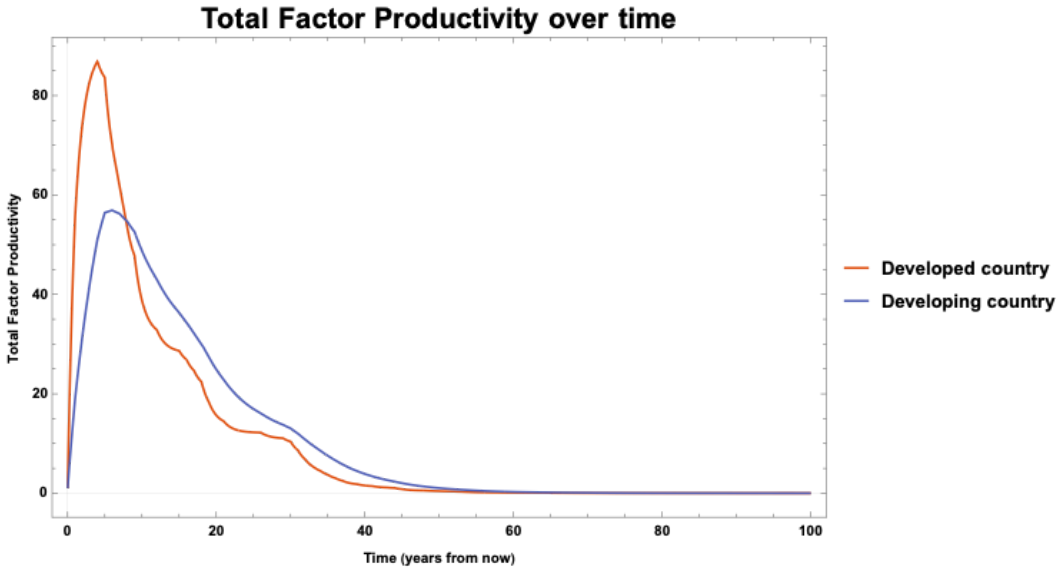

Figure 4: TFP over time for a developed and developing country.

[3] Practical 7 ECON0055, Dr Frank Witte, UCL, London, 23rd November 2023

Figure 4 tells the story of initial rapid TFP from innovations using AI in a developed country, but at 7.8 years, falls below developing country TFP. Both countries reach maximum TFP early (within 10 years). The rate of growth and decline in developing countries is slower due to a slow learning rate and its peak TFP is 30 units less than a developed country.

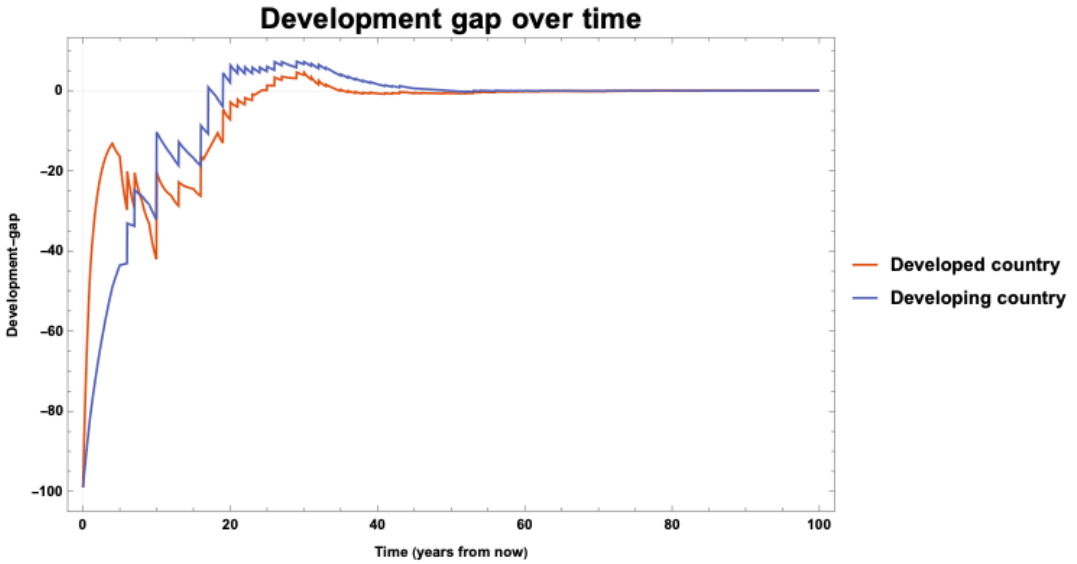

Figure 5: Development gap in technological progress over time for a developed and developing country.

Figure 5 zooms in to understand the development gap between countries. Similar to Figure 4, the developed country starts with rapid increase in technological progress until there is a hand-over at 7.8 years and progress is quicker in developing countries. The development gap in both returns to zero at around 60 years onwards.

4: Discussion

4.1: Part A - Ethical considerations

4.1.1: Social

Privacy is the biggest concern with AI, since there are not any clear boundaries as to what AI can do with personal information, such as how it acquires it and how it uses it.

Deepfakes are defined as “a video or sound recording that replaces someone's face or voice with that of someone else, in a way that appears real'' (Cambridge Dictionary, n.d.). According to a UCL study, the use of Deepfakes through audio and video impersonation was considered the most alarming aspect from AI, shown in Figure 2 (Caldwell et al., 2020). We tend to believe what our senses tell us, especially through hearing and seeing. Deepfake technologies learning capabilities can become so convincing that it is undetectable as a fake.

“96% of the Deepfake videos on the Internet are pornographic videos'' (Wang, 2019). The creation of non-consensual images and videos of real women and men has exacerbated revenge porn in an extremely dangerous way - how can we tell what is fake? Making this technology open access has been extremely detrimental, since anyone can make anything using just one photo of you. With billions of images and hours of videos being shared daily on social media, how can we sift out the fake from the real?

Deepfake detection algorithms are gaining more traction, with Kaggle creating a Deepfake Detection Challenge to promote competition and innovation of new technologies. The winner was awarded one million dollars (Kaggle, 2019). ‘Detect Fakes' is a website by MIT Media Lab (n.d.) to test yourself on how well you can spot a Deepfake. It aims to spread awareness regarding the difficulties of human detection and is being used in studies such as by Groh et al. (2021). Moreover, false positives in detecting a deepfake can be particularly dangerous, especially when analysing evidence. Depending on the context, it can lead to devastating consequences, such as social disarray. As such, awareness is the best way of educating the public what AI can do.

The Stop Killer Robots (n.d.) campaign aims to inform us of the dangers of AI and how far it can go in an extreme circumstance. It highlights the issue of digital dehumanisation and how computers cannot have the same feelings of life and existence as humans do.

‘Detroit Become Human’ (PlayStation, n.d.) is a video game which explores a dystopian world in 2038, Detroit, where androids have become superior to humans and have replaced the workforce. Playing this game highlights the extreme ‘what if’ scenario, and examines the possibility of androids gaining sentience and humanity. As you play as androids, you are exposed to discrimination, segregation and inequality. If we teach them to think and act exactly like us, are they not human? As a video game, being able to immerse yourself in the ‘other’ point of view can tell you enough about how you view AI and whether you would want it to be part of our lives.

4.1.2: Economic

An echo chamber is a community where similar beliefs are encouraged in a population, and other opinions are discouraged (Olteniceanu et al., 2022). This is similar to how an AI algorithm can be exposed to narrow perspectives in its training data, which can lead to biassed decision-making.

AI and ML have already had an impact on businesses, with OpenAI exceeding one billion dollars in revenue in 2023 alone (CNBC, 2023). This involved the introduction of ChatGPT Enterprise (OpenAI, 2023b), where companies, such as Klarna, can create a customisable GPT to the organisations needs and expectations. If a business were to input confidential information regarding a client, there can be a variety of ramifications: security risks, competitive disadvantages, damage to reputations, legal consequences and loss of trust with the client.

In attempts to leap-frog, Google has created a more advanced GPT 4 than OpenAI called Gemini (Waters, 2023). By promoting competition in the AI industry, there are prompts for quicker innovation. But will the ethics of AI be able to catch up with this? Up until December 2023, there were not any AI laws in place. On 9th December 2023, EU lawmakers agreed to a landmark bill called the AI Act (Espinoza, 2023).

4.1.3: Environmental

A study indicates that a single generative AI model can use up to 284,000 litres of water (Thomas, 2023). With the increasing demand for AI, there is more pressure to build more data centres and so the management of these requires increased water consumption. Figure 6 shows how many devices are required to run AI and each have their own life cycles that need to be sustained by energy.

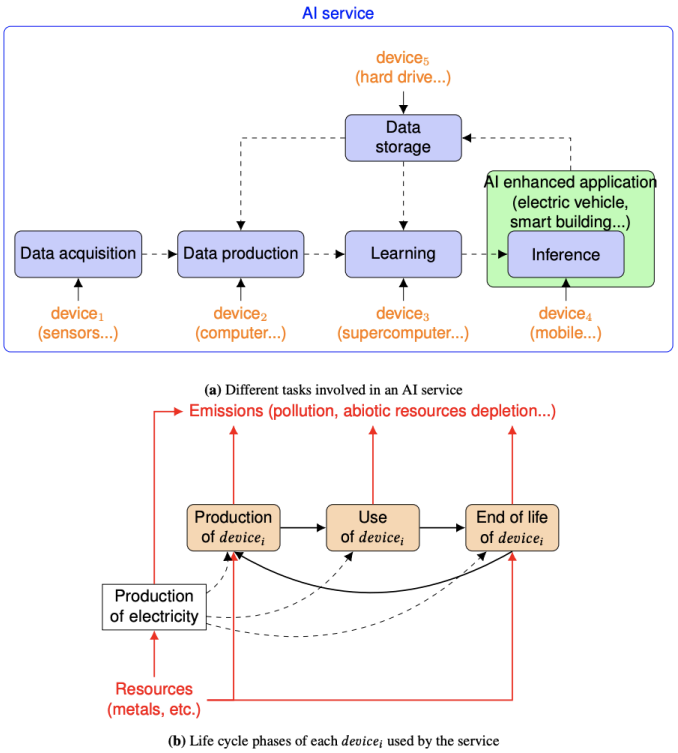

Figure 6: Lifecycle of devices used for AI services (Source: Ligozat et al., 2022).

AI and ML can be more sustainable to run through green intelligence and awareness. Firms can inform themselves using measurement tools, such as a ML emissions calculator (Schmidt et al., n.d.). Encouraging building green data centres powered by renewable energy and emphasising e-waste recycling (Kirvan, 2022) to extend the life cycle of devices used for AI (Figure 6) can be beneficial.

Generative AI can be used to create a sustainable future. Having different perspectives and prompting for different results can be efficient in innovation. It can be used as a positive feedback cycle - AI helping AI.

4.2: Part B - Influence on employment and economic growth

4.2.1: Employment

The gradient-learning model indicates that Scenario 1 leads to the most effective outcome (Figure 3). Scenario 3 does lead to a job being secured the quickest, though this means the user is well equipped with skills to use AI. Arguably, the most realistic scenario is Scenario 2. Success from using AI for job applications depends how efficiently it is being used, and since many people are new to AI, they initially would not have the skills to reap the benefits quickly . This is worsened by providing the wrong prompts for generative AI, or simply that over time, hiring firms may find alternative methods that prevent the use of AI in applications, such as timed video interviews and personality tests, thus slowing the rate of success.

Another issue is that jobs may be displaced or redefined from AI. The qualifications of someone applying may become outdated over time as AI becomes more integrated into employment (Acemoglu et al., 2022: 295). Scenario 4 in Figure 3 could represent a person taking the time to learn new skills, for example, learning programming languages such as Python for ML, and then beginning to search for jobs. It is unlikely it would take as long as shown in Figure 3, but could lead to a job being secured in less applications.

Moreover, AI has been seen as a useful tool in removing human prejudice during the recruitment process - although this depends on how it is trained since AI can learn bias without knowing what it means (Chen, 2022).

4.2.2: Economic growth

Schumpeterian growth refers to the process of innovation and creative destruction - destroying innovation to make way for new innovation (Aghion, Akcigit and Howitt, 2015). Figure 5 accounts for this by showing the fluctuations whilst reducing the development gap. After each period of reduced technological progress, it was replaced by a greater or lesser growth in the next period. The larger growth can account for imitation over innovation and technological leap-frogging.

Theoretically, we expect AI to rapidly increase TFP growth (Figure 4), though this may not be true in reality. Figure 7 indicates global TFP growth is declining and is expected to decline. AI would need to have a widespread adoption and big enough impact to overcome this decline.

Figure 7: History of TFP since 1960s (Source: Slater, 2023).

The model in Figure 4 suggests the developing country would technologically leap-frog the developed at some point. Lee (2021) examines China's successful leap-frog due to BAT corporations - Baidu, Alibaba, and Tencent - rivalling the US GAFAM corporations - Google, Apple, Facebook, Amazon, and Microsoft. However, the Thucydides Trap refers to the inevitability of war from a large power (China) displacing the one currently in power (US), which has been mirrored by the US-China trade war in 2020 (BBC News, 2020). This trap reflects the mutual decline of TFP growth in Figure 4, since Trump's tariffs attempt to discourage consumers buying imported goods from China represents him seeking to maintain economic dominance.

This model is less accurate for underdeveloped economies according to Rimmer (1961), since there is a dependency on foreign technology which would require a large leap-frog innovation to overcome. With brain drain leading to high quality human capital moving out of underdeveloped countries, the likelihood of this happening is greatly reduced.

5: Conclusion

This essay has explored the moral implications that AI and ML may have going forward, and addressed the importance of setting boundaries and ethical considerations to prevent a worse-case scenario from occurring.

Why should ethical considerations be integrated into the design and development process of AI and ML systems?
The social, economic, and environmental concerns behind AI are attributed to uncertainty. Privacy and exploitation, anti-competitiveness, and unsustainable practices can be resolved through transparency and awareness, legislation, and green technology, respectively.

How do AI and ML influence employment and economic growth?
The gradient-learning model implies using AI for job applications leads to a better chance of securing a job, though this can be mitigated by hiring firms redefining skills required for a job to account for the inclusion of AI in society.

The Schumpeterian economic growth model highlights the impact of AI on TFP in a developed and developing country, but suggests in the long-run, it may not be as effective since innovation can be inhibited through the Thucydides Trap.

This essay examines the inevitable impact of AI on our lives and careers, but to what scale will it actually affect us?

6: Reference List

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, 40(S1), pp.S293–S340. doi:https://doi.org/10.1086/718327.

Aghion, P., Akcigit, U. and Howitt, P. (2015). The Schumpeterian Growth Paradigm. *Annual Review of Economics*, 7(1), pp.557–575. doi:https://doi.org/10.1146/annurev-economics-080614-115412.

BBC News (2020). A quick guide to the US-China trade war. *BBC News*. [online] 16 Jan. Available at: https://www.bbc.co.uk/news/business-45899310.

Caldwell, M., Andrews, J.T.A., Tanay, T. and Griffin, L.D. (2020). AI-enabled future crime. *Crime Science*, 9(1). doi:https://doi.org/10.1186/s40163-020-00123-8.

Cambridge Dictionary (n.d.). *deepfake*. [online] dictionary.cambridge.org. Available at: https://dictionary.cambridge.org/dictionary/english/deepfake.

Chen, Z. (2022). Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. *Cognition, Technology & Work*, [online] 25(1), pp.135–149. doi:https://doi.org/10.1007/s10111-022-00716-0.

CNBC (2023). *OpenAI on pace for $1.3 billion in 2023 revenue, up 4543% from last year*. [online] CNBC. Available at: https://www.cnbc.com/video/2023/10/13/techcheck-openai-on-pace-for-1-point-3-billion-in-2023-revenue-up-4543-percent-from-last-year.html [Accessed 12 Dec. 2023].

della Briotta Parolo, P., Kujala, R., Kaski, K. and Kivelä, M. (2020). Tracking the cumulative knowledge spreading in a comprehensive citation network. *Physical Review Research*, 2(1). doi:https://doi.org/10.1103/physrevresearch.2.013181.

Dictador (n.d.). *Dictador*. [online] dictador.com. Available at: https://dictador.com [Accessed 12 Dec. 2023].

Espinoza, J. (2023). *EU agrees landmark rules on artificial intelligence*. [online] Financial Times. Available at: https://www.ft.com/content/d5bec462-d948-4437-aab1-e6505031a303.

Groh, M., Epstein, Z., Firestone, C. and Picard, R. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1). doi:https://doi.org/10.1073/pnas.2110013119.

Kaggle (2019). *Deepfake Detection Challenge*. [online] www.kaggle.com. Available at: https://www.kaggle.com/c/deepfake-detection-challenge/overview.

Kirvan, P. (2022). *What is a green data center?* [online] SearchDataCenter. Available at: https://www.techtarget.com/searchdatacenter/definition/green-data-center.

Lee, K. (2021). *China's Technological Leapfrogging and Economic Catch-up*. Oxford University Press. doi:https://doi.org/10.1093/oso/9780192847560.001.0001.

Ligozat, A.-L., Lefevre, J., Bugeau, A. and Combaz, J. (2022). Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. *Sustainability*, [online] 14(9), p.5172. doi:https://doi.org/10.3390/su14095172.

MIT Media Lab (n.d.). *DeepFakes, Can You Spot Them?* [online] detectfakes.media.mit.edu. Available at: https://detectfakes.media.mit.edu.

Olteniceanu, A.-M., Van Bevervoorde, M.E., Flos, M. and Torlaini, R. (2022). *The Network Science of Echo Chambers and Why It Matters – The Network Pages*. [online] The Network Pages. Available at: https://networkpages.nl/the-network-science-of-echo-chambers-and-why-it-matters/ [Accessed 12 Dec. 2023].

OpenAI (2023a). *ChatGPT*. [online] chat.openai.com. Available at: https://chat.openai.com/auth/login.

OpenAI (2023b). *Introducing ChatGPT Enterprise*. [online] openai.com. Available at: https://openai.com/blog/introducing-chatgpt-enterprise.

PlayStation (n.d.). *Detroit: Become Human*. [online] store.playstation.com. Available at: https://store.playstation.com/en-gb/product/EP9000-CUSA08308_00-DETROIT000000001 [Accessed 12 Dec. 2023].

Radicchi, F., Fortunato, S. and Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, 105(45), pp.17268–17272. doi:https://doi.org/10.1073/pnas.0806977105.

Rimmer, D. (1961). Schumpeter and the Underdeveloped Countries. *The Quarterly Journal of Economics*, 75(3), p.422. doi:https://doi.org/10.2307/1885132.

Schmidt, V., Luccioni, A.S., Lacoste, A. and Dandres, T. (n.d.). *Machine Learning CO2 Impact Calculator*. [online] mlco2.github.io. Available at: https://mlco2.github.io/impact/#compute.

Slater, A. (2023). *AI and productivity growth – a note of caution*. [online] *Oxford Economics*. Available at: https://www.oxfordeconomics.com/wp-content/uploads/2023/10/AI-and-productivity-growth-a-note-of-caution.pdf?pi_content=d3ba92741e77f9487acb2b5d3d2d9725df3515380ea4e45ce35c9e1d4a5df145 [Accessed 15 Dec. 2023].

Stop Killer Robots (n.d.). *Stop Killer Robots*. [online] Stop Killer Robots. Available at: https://www.stopkillerrobots.org/stop-killer-robots/.

Thomas, S. (2023). *AI revolution: what's the environmental impact?* [online] Schroders. Available at: https://www.schroders.com/en-gb/uk/intermediary/insights/ai-revolution-what-s-the-environmental-impact-/.

Wang, C. (2019). *Deepfakes, Revenge Porn, And The Impact On Women*. [online] Forbes. Available at: https://www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and-the-impact-on-women/.

Waters, R. (2023). *Google pins its hopes on Gemini to leapfrog GPT-4*. [online] Financial Times. Available at: https://www.ft.com/content/e5cc4e36-efe3-4491-b435-75a712533257 [Accessed 12 Dec. 2023].

Yoon, J. (2020). Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Computational Economics*, 57(1). doi:https://doi.org/10.1007/s10614-020-10054-w.