UK AI Policy: An Engineering and Safety Perspective

Joe Ingham and Noah Cadwaladr Evans

Department of Engineering, University of Cambridge, Cambridge, CB2 1TN, United Kingdom.

*Corresponding author(s). E-mail(s): ji291@cam.ac.uk;
Contributing authors: enc26@cam.ac.uk;

Abstract

As the world accepts AI into everyday life, the UK aims to become a world leader in AI policy and implementation, boosting the national AI industry, and with this, the economy. While sounding promising, this needs to be done with the utmost care to ensure that the public is not exploited in the process. This manuscript critically reviews the UK AI policy from an engineering and safety perspective, highlighting hopeful areas, ethical concerns, and areas for improvement.

Keywords: Artificial Intelligence, AI Ethics, UK Science Policy, UK Industry, Opinion

1 Science → Policy

The rapid development and integration of Artificial Intelligence into professional and academic life provides a new opportunity for the UK government to refresh it's struggling industry. However, it is important that AI is adopted in a conscientious manner, one that will benefit the public, not harm them. To do this, care must be taken when implementing AI policy and regulation.

2 Introduction

In a world that is continuously generating more data and rapidly integrating Artificial Intelligence (AI) into every application, the UK government has decided that the best approach is not to ride the AI wave, but to generate it. With forethought and meticulous planning, the UK government believes that it can become an "AI superpower",

leading the world in research and adoption. The envisioned goal is a proverbial golden age for the nation's technology market resulting in newfound economic prosperity.

As a world leader, the UK would be a major exporter of software that is increasingly utilized in every facet of life. As the global stage is being drawn, advancements are being announced weekly, if not daily. Two big hitters in the AI playing field, as of writing, are Nvidia and OpenAI, both American, and both highly valued. As a testament to the ever-changing landscape, barely weeks after the UK announced its AI plan, the Chinese company "DeepSeek" released their flagship AI model, causing shock waves in the market [1]. Although DeepSeek's announcement resulted in Nvidia's stock value reducing by approximately 10%, the latter is still worth an astonishing amount and at points has been worth more than the entire market cap of the UK [2].

Although the UK is already home to AI powerhouses Wayve and Deepmind, the government wants to supercharge the country's AI industry, hoping to gain a firmer foothold in the tumultuous and ever-changing field. If the government can promote the creation of an Nvidia-level player, it could dramatically increase the tax revenue the UK generates.

Investment in AI industry in the UK is continuing to grow [3]; to capitalise on this the UK government commissioned an "AI Opportunities Action Plan" [4] from Matt Clifford CBE, a leading AI entrepreneur. This document acts as a set of recommendations with the target of invigorating the AI industry in the UK. The government's response highlights the main steps it has said it will take towards "laying the foundations to enable AI" [5].

These foundations can be distilled into 6 main sections: Building AI infrastructure to help support the expansion of AI in the UK; creating "AI Growth Zones", which are areas with promoted investment opportunities via easier access to planning approval; the creation of an AI Energy Council, to address the AI power crisis; the creation of a "National Data Library" with the aim of providing public data for creation of new AI models; the introduction of new AI skills and skills capacity via training and the attraction of new talent; and underpinning all of these is an investigation on "efficient" regulation.

3 Discussion

3.1 Engineering Research Impact

The plan states an intention to increase UK computing capacity "by at least 20X" within the next five years. The UK already has an extensive range of supercomputers that are predominantly used for research [6]. However, this is evidently insufficient for the envisioned development and implementation of AI.

The increased availability will allow larger models to be created which can have an improved "understanding" of the data they are provided with; this is particularly useful in engineering as it enables more detailed analysis of complex problems. This could lead to AI-enabled analysis and reviews of design (although care needs to be taken), reducing the workload for the average person. This contrasts with existing analytical tools frequently used in design practice which reduce the burden on the engineer, how much can an AI tool improve upon the engineers current modus operandi?

More complex models also lead to more "human-like" reasoning capabilities and an increased perceived intelligence. This increase in intelligence would be beneficial for safety-critical applications, such as healthcare or driving, but can an AI tool ever be left with ultimate responsibility?

3.2 Power Supply and Consumption

Powering AI data centres is a major concern, with carbon emissions skyrocketing as companies adopt and research AI solutions. Google's greenhouse emissions increased a staggering 48% as a direct result of AI's power guzzling requirements [7].

To combat this, the UK government has announced the creation of a group consisting of industry leaders in AI and energy, dubbed the "AI Energy Council". The council intends to consult on power requirements and how to address this with "renewable and innovative energy solutions".

Of particular note is the mention of Small Modular Reactors (SMRs), compact nuclear reactors designed to be compatible with each other such that they can be combined when required.

Although public opinion towards nuclear power is improving, it is still low [8][9]. The UK not only needs to convince the public that nuclear is safe, but that it is the most viable solution for our energy needs. The development of SMRs will help alleviate climate change and our dependence on foreign power imports (without falling into the trap Germany did when they closed their nuclear reactors [10]). It also has the potential to kickstart the UK's exceptionally slow energy infrastructure development (think Hinkley Point C [11][12]).

As the world looks towards sustainable infrastructure, if the UK's focus on AI also advances SMR research, it has the potential to become a world leader in this area as well, putting it in a strong position as the rest of the world begins to adopt both of these technologies.

3.3 AI Equality and Ethicality

As AI adoption grows, it is important that it is implemented ethically and with the benefit of the general public in mind. AI should be a tool for public good, especially if it is funded by the taxpayer. Advancements in healthcare and other similar target areas should be prioritised over small conveniences such as ChatGPT or Google Gemini.

It is important to note that when using AI it can be hard to assign accountability. To quote IBM, "A computer can never be held accountable, therefore a computer must never make a management decision". When a computer makes a mistake, who is held responsible? The programmer? The manager? Or the government for incentivising the AI adoption? This holds true for any sector that utilises AI, whether education ([13]), healthcare ([14][15]) or elsewhere [16][17].

If an AI is making decisions, it also has the potential to perpetuate any biases that are prevalent in the data used to create the AI model. As such AI researchers need to be aware of such risks and ensure they are as minimal as possible. Care has to be taken to ensure that underlying data biases are removed, otherwise, the AI risks

preserving racism ([18][19][20][21]), sexism ([22][23][24][25]), or other such social biases and prejudices [26][27][28][29].

Alongside data prejudices, it is also important to reduce negative physical effects of AI. AI Growth zones have the potential to revitalise old UK industrial areas by bringing new industry into locations where industry has been long since abandoned. However, it has to be ensured that the locals are not priced out of their own homes when the new-wave of tech entrepreneurs move to the growth zones. (i.e. gentrification [30][31][32]) Could AI training have priority for those areas to give individuals the chance to stay up to speed with the change around them?

To ensure AI fairness and equality, there have to be strong regulations in place, to help guide those utilising and researching AI, and to protect the public from malicious AI actors.

3.4 Governance and Safety

Overall, one of the most important aspects of this governmental policy, will be the regulatory effects. With proper regulation, the government can ensure that AI is being used responsibly and safely.

However, the policy outline discusses slashing regulation in key areas to "drive innovation". This is risky, particularly for engineering AI is seeing increased usage in design and verification processes. Although this may decrease the workload for engineers, there needs to be accountability and trust in the designs, that is lacking when generated by faceless computers.

This rings true in other safety-critical fields such as health and energy, if things can go wrong and lives are at risk, there has to be responsibility, otherwise nothing is stopping companies from using and blaming AI for when something (inevitably) goes wrong.

In general, AI has other well-established risks, especially generative AI, including deep-faking whether for scams or for creating "non-consensual intimate imagery" (NCII)[33]. Although we are not concerned with the idea of AI taking over the world "terminator-style", we are concerned with those less educated on AI, those who are more likely to become victim to people who use it for harm. To combat this, there need to clear laws in place, to deter and punish those using AI for socially nefarious purposes. For example, it is entirely possible (and a good idea) to make generating NCIIs an offence that also places you on the sex offender registry. This is a step the government has stated they intend to make [34][35]. The public would also benefit from the government releasing resources that highlight new social risks from AI, akin to resources about scams and cyber-security.

The EU commission has also recently released guidelines on prohibited AI use in the EU, it is helpful in identifying harmful uses of AI, such as fraud [36]. Will the UK follow suit and highlight areas of illegal AI activity?

To ensure safety, the government and private sector should be completely transparent on what AI is being used for and how they are going to regulate it. This way, it allows the public body and AI experts to decide whether they think the direction of AI in the UK is a safe and beneficial one.

It is hard to tell what kind of regulations will be in place or removed as the government has stated that DSIT (Department for Science, Innovation, and Technology) will release reports over the coming year, detailing what action will be taken. The government is currently legislating on the usage of copyright materials by or for AI and, worryingly, is suggesting that AI trainers should have access to copyright materials, without express permission of the creators [37][38]. This will be incredibly harmful for creators and artists and will simply legalise what some big AI companies are already doing [39][40].

4 Conclusion

Overall, the policy presents opportunities for researchers in a breadth of fields, both for those researching foundational AI theory and for those utilising AI for data science in their own research. The government has indicated that the AI revolution will be sustainably built and powered, access to data for researchers will be greatly improved, and the burden of menial tasks for the working population will be significantly reduced. On a surface level, it looks very promising. However, a lack of clear regulatory statements is concerning. To ensure that the policy is ethically enacted, creating regulation that protects safety-critical areas and the public's rights should be at the forefront; it must be ensured that people's privacy, health, and livelihoods are not at risk. AI should be used to make everybody's lives easier, not more complicated. When the UK is attempting to generate its own AI boom, it cannot be at the expense of the public body. This plea to be considerate and not rash is in contrast with the extreme pace at which this area is developing, the pace at which the government and industry in the UK must keep up with if the goals set out by the AI action plan are to be realised. But without careful consideration, an under-regulated AI industry may do more damage than good.

References

[1] Why Nvidia investors are spooked by Chinese AI upstart DeepSeek — ft.com. https://www.ft.com/content/ee83c24c-9099-42a4-85c9-165e7af35105. [Accessed 28-01-2025]

[2] Nvidia is worth more than the entire French and British stock markets — sherwood.news. https://sherwood.news/markets/nvidia-valuation-french-british-stock-market-capitalization-comparisons/. [Accessed 28-01-2025]

[3] State of UK Deep Tech 2024 — Read the report — enterprisehub.raeng.org.uk. https://enterprisehub.raeng.org.uk/deep-tech-2024. [Accessed 28-01-2025]

[4] AI Opportunities Action Plan — gov.uk. https://www.gov.uk/government/publications/ai-opportunities-action-plan. [Accessed 28-01-2025]

[5] AI Opportunities Action Plan: government response — gov.uk. https://www.gov.uk/government/publications/ai-opportunities-action-plan-government-response/ai-opportunities-action-plan-government-response. [Accessed 28-01-2025]

[6] Facilities — hpc-uk.ac.uk. https://www.hpc-uk.ac.uk/facilities/. [Accessed 28-01-2025]

[7] https://www.facebook.com/bbcnews: AI means Google's greenhouse gas emissions up 48https://www.bbc.co.uk/news/articles/c51yvz51k2xo. [Accessed 23-01-2025]

[8] Public Attitudes to Nuclear Power. Nuclear Energy Agency/Organisation for Economic Co-operation and Development, Paris (2010). OCLC: 1039240590

[9] DESNZ Public Attitudes Tracker: Energy infrastructure and energy security, Spring 2024, UK — gov.uk. https://www.gov.uk/government/statistics/desnz-public-attitudes-tracker-spring-2024/desnz-public-attitudes-tracker-energy-infrastructure-and-energy-security-spring-2024-uk. [Accessed 23-01-2025]

[10] After scrapping nuclear reactors, Germany to spend billions on new gas power plants — politico.eu. https://www.politico.eu/article/nuclear-reactors-germany-invest-gas-power-plants-energy/. [Accessed 23-01-2025]

[11] https://www.facebook.com/WorldNuclearNews: EDF announces Hinkley Point C delay and rise in project cost — world-nuclear-news.org. https://www.world-nuclear-news.org/articles/edf-announces-hinkley-point-c-delay-and-big-rise-i. [Accessed 31-01-2025]

[12] https://www.theguardian.com/profile/jillian-ambrose: Hinkley Point C owner warns fish protection row may further delay nuclear plant — theguardian.com. https://www.theguardian.com/business/2025/jan/30/hinkley-point-c-owner-warns-fish-row-may-further-delay-nuclear-plant. [Accessed 31-01-2025]

[13] Klimova, B., Pikhart, M., Kacetl, J.: Ethical issues of the use of AI-driven mobile apps for education. Frontiers in Public Health **10**, 1118116 (2023) https://doi.org/10.3389/fpubh.2022.1118116 . Accessed 2025-01-23

[14] Habli, I., Lawton, T., Porter, Z.: Artificial intelligence in health care: accountability and safety. Bulletin of the World Health Organization **98**(4), 251–256 (2020) https://doi.org/10.2471/BLT.19.237487 . Accessed 2025-02-28

[15] Smith, H.: Clinical AI: opacity, accountability, responsibility and liability. AI & SOCIETY **36**(2), 535–545 (2021) https://doi.org/10.1007/s00146-020-01019-6 .

Accessed 2025-02-28

[16] Raja, A.K., Zhou, J.: AI Accountability: Approaches, Affecting Factors, and Challenges. Computer **56**(4), 61–70 (2023) https://doi.org/10.1109/MC.2023.3238390 . Accessed 2025-02-28

[17] Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., Weinberger, D., Weller, A., Wood, A.: Accountability of AI Under the Law: The Role of Explanation. arXiv. Version Number: 3 (2017). https://doi.org/10.48550/ARXIV.1711.01134 . https://arxiv.org/abs/1711.01134 Accessed 2025-02-28

[18] Fountain, J.E.: The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. Government Information Quarterly **39**(2), 101645 (2022) https://doi.org/10.1016/j.giq.2021.101645 . Accessed 2025-02-28

[19] Tian, J., Xie, H., Hu, S., Liu, J.: Multidimensional Face Representation in a Deep Convolutional Neural Network Reveals the Mechanism Underlying AI Racism. Frontiers in Computational Neuroscience **15**, 620281 (2021) https://doi.org/10.3389/fncom.2021.620281 . Accessed 2025-02-28

[20] Hong, J.-W., Williams, D.: Racism, responsibility and autonomy in HCI: Testing perceptions of an AI agent. Computers in Human Behavior **100**, 79–84 (2019) https://doi.org/10.1016/j.chb.2019.06.012 . Accessed 2025-02-28

[21] Lawrence, C.: Hidden in White Sight: How AI Empowers and Deepens Systemic Racism, 1st edn. Chapman and Hall/CRC, Boca Raton (2023). https://doi.org/10.1201/9781003368755 . https://www.taylorfrancis.com/books/9781003368755 Accessed 2025-02-28

[22] Brown, L.M.: Gendered Artificial Intelligence in Libraries: Opportunities to Deconstruct Sexism and Gender Binarism. Journal of Library Administration **62**(1), 19–30 (2022) https://doi.org/10.1080/01930826.2021.2006979 . Accessed 2025-02-28

[23] Elder, A.: Siri, Stereotypes, and the Mechanics of Sexism. Feminist Philosophy Quarterly **8**(3/4) (2022) https://doi.org/10.5206/fpq/2022.3/4.14294 . Accessed 2025-02-28

[24] Hou, T.-Y., Tseng, Y.-C., Yuan, C.W.T.: Is this AI sexist? The effects of a biased AI's anthropomorphic appearance and explainability on users' bias perceptions and trust. International Journal of Information Management **76**, 102775 (2024) https://doi.org/10.1016/j.ijinfomgt.2024.102775 . Accessed 2025-02-28

[25] Hong, J.-W., Choi, S., Williams, D.: Sexist AI: An Experiment Integrating CASA

and ELM. International Journal of Human–Computer Interaction **36**(20), 1928–1941 (2020) https://doi.org/10.1080/10447318.2020.1801226 . Accessed 2025-02-28

[26] İLERİ, S.Ç.: Gender bias in ai: How to dismantle prejudices. ENGLISH STUDIES A MULTIFACETED LENS, 365

[27] Bias, A.I.: Systematic prejudices (2024)

[28] Joisten, K., Renner, T.: Technological biases and human prejudices: A technoethical challenge. In: 2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 01–06 (2024). IEEE

[29] Baines, A., Gruia, L., Collyer-Hoar, G., Rubegni, E.: Playgrounds and prejudices: Exploring biases in generative ai for children. In: Proceedings of the 23rd Annual ACM Interaction Design and Children Conference, pp. 839–843 (2024)

[30] Stumer, C.R.: Super gentrification: The influence of silicon valley in san francisco. Paper for Geography **457** (2014)

[31] Opillard, F.: Resisting the Politics of Displacement in the San Francisco Bay Area: Anti-gentrification Activism in the Tech Boom 2.0. European journal of American studies **10**(3) (2015) https://doi.org/10.4000/ejas.11322 . Accessed 2025-01-30

[32] Stehlin, J.: The Post-Industrial "Shop Floor": Emerging Forms of Gentrification in San Francisco's Innovation Economy. Antipode **48**(2), 474–493 (2016) https://doi.org/10.1111/anti.12199 . Accessed 2025-01-30

[33] Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., Choi, Y., Fox, P., Garfinkel, B., Goldfarb, D., Heidari, H., Ho, A., Kapoor, S., Khalatbari, L., Longpre, S., Manning, S., Mavroudis, V., Mazeika, M., Michael, J., Newman, J., Ng, K.Y., Okolo, C.T., Raji, D., Sastry, G., Seger, E., Skeadas, T., South, T., Strubell, E., Tramèr, F., Velasco, L., Wheeler, N., Acemoglu, D., Adekanmbi, O., Dalrymple, D., Dietterich, T.G., Felten, E.W., Fung, P., Gourinchas, P.-O., Heintz, F., Hinton, G., Jennings, N., Krause, A., Leavy, S., Liang, P., Ludermir, T., Marda, V., Margetts, H., McDermid, J., Munga, J., Narayanan, A., Nelson, A., Neppel, C., Oh, A., Ramchurn, G., Russell, S., Schaake, M., Schölkopf, B., Song, D., Soto, A., Tiedrich, L., Varoquaux, G., Yao, A., Zhang, Y.-Q., Albalawi, F., Alserkal, M., Ajala, O., Avrin, G., Busch, C., Carvalho, A.C.P.d.L.F., Fox, B., Gill, A.S., Hatip, A.H., Heikkilä, J., Jolly, G., Katzir, Z., Kitano, H., Krüger, A., Johnson, C., Khan, S.M., Lee, K.M., Ligot, D.V., Molchanovskyi, O., Monti, A., Mwamanzi, N., Nemer, M., Oliver, N., Portillo, J.R.L., Ravindran, B., Rivera, R.P., Riza, H., Rugege, C., Seoighe, C., Sheehan, J., Sheikh, H., Wong, D., Zeng, Y.: International AI Safety Report. arXiv. Version Number: 1 (2025). https://doi.org/10.48550/ARXIV.2501.17805 . https://arxiv.org/abs/2501.17805 Accessed 2025-01-31

[34] Nowell, T.: New laws criminalise the sharing of intimate deepfakes without consent — refuge.org.uk. https://refuge.org.uk/news/new-laws-criminalise-the-sharing-of-intimate-deepfakes-without-consent/. [Accessed 31-01-2025]

[35] MP, A.D.J.: Tackling Intimate Image Abuse and Sexually Explicit Deepfakes. https://questions-statements.parliament.uk/written-statements/detail/2025-01-07/hcws354. [Accessed 31-01-2025]

[36] Commission publishes the Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act. — digital-strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act. [Accessed 10-02-2025]

[37] Media, P.: Lords amendments offer AI copyright protection for publishers — pressgazette.co.uk. https://pressgazette.co.uk/news/government-defeated-lords-copyright-ai-scraping/. [Accessed 11-02-2025]

[38] UK Government proposes copyright and AI reform mirroring EU approach — linklaters.com. https://www.linklaters.com/en/insights/blogs/digilinks/2025/january/uk-government-proposes-copyright-and-ai-reform-mirroring-eu-approach. [Accessed 11-02-2025]

[39] Meta staff torrented nearly 82TB of pirated books for AI training — court records reveal copyright violations — tomshardware.com. https://www.tomshardware.com/tech-industry/artificial-intelligence/meta-staff-torrented-nearly-82tb-of-pirated-books-for-ai-training-court-records-reveal-copyright-violations. [Accessed 11-02-2025]

[40] Pillay, T.: Timeline of Recent Accusations Leveled at OpenAI, Sam Altman — time.com. https://time.com/6986711/openai-sam-altman-accusations-controversies-timeline/. [Accessed 11-02-2025]