The Next Generation of AI Tools are Coming, do we Need to Better Regulate Them?

Michael Aubrey
University of South Wales, Blackwood, UK

04184025@students.southwales.ac.uk

Abstract: This research investigates the importance for robust regulation and legislation to properly govern the development of 'Frontier Artificial Intelligence Systems'. To do this a comparison of approaches of both existing and proposed legislation has been taken which includes from the US, UK and EU. This has involved reading both summations, assessments and opinions from both academic writers and the media and making comparisons with other industry regulations issues such as Social Media. The key findings of this research highlight the different approaches being taken to the same problem. Legislation that came into force on August the 1st 2024 from the EU takes a safety-first approach, identifying risk levels from 'unacceptable risk' that would be prohibited, down to 'minimal risk' which would remain unregulated. This is compared to the UK White Paper, which advocated an innovation first approach with a secondary focus on safety. With the potential risks associated with AI enhanced cyber-attacks and the spread of disinformation across different platforms, this research emphasises the importance of strong regulation in relation to safety and ensuring this happens from the outset in the development of 'Frontier AI' and is not an afterthought.

Keywords: Artificial intelligence, Cambridge analytica, Disinformation, Governance, Regulation, Safety

1. Introduction

The purpose of this research is to try to identify the current situation in relation to the governance and regulation of this industry in areas of global significance to the development of AI and compare their approaches. Then to investigate some of the potential issues that current AI technology creates and to speculate on how more powerful AI systems might accentuate these using examples that the currently unregulated social media market provides and to make some recommendations as to how governance and regulation may be improved based on the conclusions provided. It is a helpful to look at some of the lessons that we can maybe learn from the social media governance or lack of and to identify the issues and concerns in that area, as we can see integration from AI platforms such as 'Grok' into highly potent and influential Social Media platforms such as X. With social media being a relatively new phenomenon and a fast-evolving technology therefore using the experience with it will help us to gauge where we may see potential future issues from the progression of AI, its governance and regulation (Chauhan, Krishna Deo Singh. 2023).

There are numerous stories that exist in the global media extolling the virtues of Generative AI (Gen AI) and giving examples of its positive impact on, for example, education, where it can be used to improve efficiencies of staff and students in terms of planning and structuring lesson content and assignment work. This is of course a positive and using this type of technology for this type of purpose should be encouraged. However, we also see the negative where Gen AI is used to effectively cheat and claim credit for work that is not written, either by the student or by the academic but by the AI, and at some point, we are going to reach a point where either the positive outweighs the negative or vice versa, and what will this mean for trust, safety, data protection and the potential to spread misinformation and influence?

The expectations of this research are that it will likely show a significant area of concern around regulation especially when considering commercial interests. It is likely that significant lobbying has already taken place and will continue to do so, to attempt to ensure that there is a lighter touch to regulation in the development of these systems and ensuring the interests of technology companies is firmly represented. As frontier AI is likely to be a global phenomenon, as social media was back in the early part of the 21st century, it is likely that there will be significant global interest in harnessing the technology as quickly as possible. Frontier AI is described as an umbrella term for models of AI that outperform existing models in what they are capable of or task that they are able to perform (Jaquemard, 2023). This has been a long-standing goal for many developers of AI systems such as OpenAI, the company behind generative AI products such as ChatGPT. The current generation of AI tools already have the potential to cause huge problems throughout the world the question that we need to ask is will so called Frontier or General AI exasperate these problems? This is a complex question and in examining answers we need to first identify some of the potential issues that exist and how, through legislation and regulation, we might try to address these issues.

2. Review of Available Literature

Due to the ever-evolving nature of the technology behind AI and coupled with the fact this is a new field of study makes finding relevant academic sources in relation to the regulation of AI difficult especially as some of that regulation has not yet come to pass. However, we can use an existing technology like Social-Media (SM) and explore the approaches to that and how some of these lessons maybe learned in regulating AI development.

In order to see the effects of a perceived lack of regulation we only need to look as far as the SM market. In their article, Nyabola discusses at length the effect that social media has had on democracy and the challenges for regulating it. In its earliest conception SM was greeted with optimism and was seen as a great opportunity to build global connectivity, however has massive potential to undermine democracy in a number of ways for example the reinforcement of insular and damaging perspectives, appealing users inbuilt prejudices and the ability of bad actors through financial measures to buy greater visibility and influence and by exploiting a network that was developed with the best intentions (for example X) found a way to spread considerable misinformation and like we are now seeing with AI this creates a considerable regulatory challenge that as of yet has not been sufficiently addressed. Before the US election of 2016 the effect of SM in politics was seen as a complex issue, however after the election and the controversies that came with the influence SM had on the election of Donald Trump the challenge was seen as a global one rather that something that just affected less stable democracies (Nyabola, 2023).

We may ask, should we be concerned with the regulation of these industries, so to demonstrate the problem, we can point to examples such as the Cambridge Analytica (CA) scandal (Matara, M. et al. 2022). In 2016 two of the world's top five economic powers held significant democratic events in June the UK population were asked to vote in a referendum to decide the UK's membership of the European Union (EU) and in the US the presidential election was to take place in November. Ultimately these events produced results that had repercussions across the world, Brexit happened and of course Donald Trump was elected president. It later came to light that these outcomes may have been influenced by the collection of personal data through an app on Facebook called "my digital life". CA was ultimately able to gain access to data from it is estimated 87 million Facebook users. Even though the app only had an estimated 320,000 users, data was able to be harvested from not only these app users but also their friends and associates who they interacted with over Facebook, who crucially were unaware and had not consented to their data being used in this way. This allowed CA to create very accurate voter profiles and predict their voting intention this enabled them to better target individuals with advertisements that could influence an individual voter to vote the way CA or CA's clients wanted them too (Fadi Safieddine and Ibrahim, 2020), this was described effectively as a weaponisation of AI.

Shortly after the scandal broke the European Parliament made a resolution that recognised that the scandal had significant implications and the potential to manipulate past and future elections and referenda. In the UK, the House of Commons in 2019 issued a damaging report noting the influence of Russia on disinformation campaigns in 2016 that can be linked directly to the result of the Brexit referendum. The CA scandal bought to light fundamental weaknesses in existing legislation most notably GDPR regulation 679/216 specifies the right to protection of personal data, articles 7 and 8 of the Charter of Fundamental Rights and the EU-US privacy shield agreement of 2016 that sought to allow the easy transfer of data from the EU to certified companies in the US. After the scandal the privacy shield was no longer seen as a valid instrument for this purpose (Buruiană, Andreea, 2021). Whilst the weaknesses have certainly been highlighted adequate solutions have yet to present themselves, this will be partly down to the nature of legal challenges both to and from companies like 'Meta' and the fact that these companies were able to act with impunity with very little in the way of consequence highlights their global influence but also the significant weakness in legislation and regulation.

After the CA scandal it emerged that peoples trust in SM companies to keep their data secure was significantly compromised, Figure 1 shows we are seeing an erosion of trust in SM companies globally with only Governments being shown as less trustworthy as show in the Ipsos Global Trust monitoring report published in 2023. The figures in the report do indicate that trust is improving in all sectors however SM are still amongst the lowest scorers in this report.

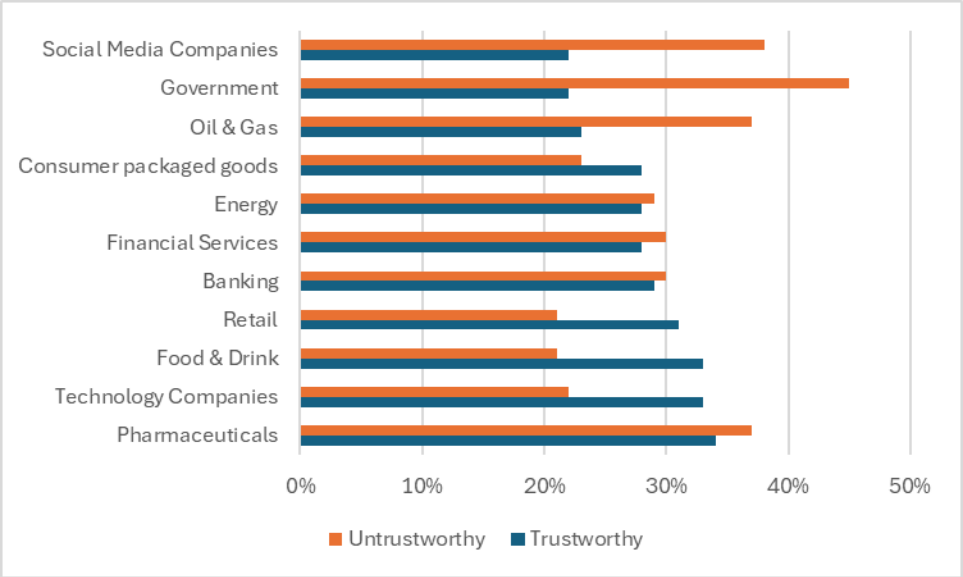

Figure 1: Global average levels of trust by sector in 2022/23 - (Ipsos, 2022)

What is interesting is the report also publishes the global attitude to regulation or lack of it, and once again SM companies score significantly high as shown in Figure 2 in terms of too little regulation.

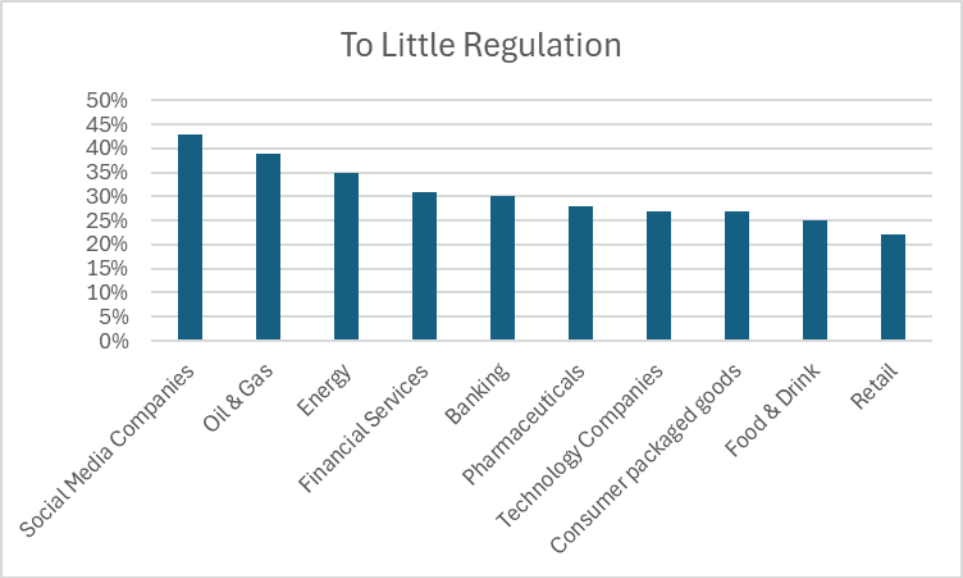

Figure 2: Global attitude toward regulation - (Ipsos, 2022)

These numbers are valuable for a number of reasons but in particular they show the value placed on trust and regulation. Although we would like more regulation, the irony is we do not seem to place trust in the people responsible for bringing about and enforcing that regulation i.e., Governments, as highlighted in Figure 1.

Unlike SM generative AI allows content that is no longer exclusively user generated but synthetic, and generated with great speed. Therefore, we can foresee many potential issues on the horizon where generative AI tools are used to created fake content or misinformation that is maybe not so damaging in isolation but when combined with the power of the social networks with multiple delivery channels this becomes a very different problem. In his article James Pinell makes this point and also espouses the need for parliaments around the world to act, also noting that;

> *This is as "bad" as these technologies are going to be. Artificial Intelligence is going to get more capable, not to mention more usable, over time. (Pinell, 2024).*

Whilst we may see Gen AI applications as powerful and advanced in their current form, they are only going to become more effective at what they do. This is a strong argument for governments to get a grip on this at the

earliest possible phase. In response to this the Commonwealth Parliamentary Association (CPA) have produced a handbook, designed to educate users on how to spot disinformation especially from synthetic generated sources that can be extremely difficult. This handbook contains strategies for combatting disinformation (Pinell, 2024), and whilst this is all a very welcome recognition of the problem and may go some way to help alleviate it, these actions can be simply ignored by bad actors who wish to spread this type of content.

3. Research Methodology

To substantiate the conclusions and recommendations made in this research a secondary methodology has been adopted using existing data and research from credible journalistic and academic sources. The scope of the paper brings in sources within a four-year time frame from 2020 although there is reference to work done from as far back as 2016. Geographically the focus is primarily on the United Kingdom (UK) and Europe however where pertinent references are made to China, the United States (US) and Australia. In searching various academic databases, although some work has been done in this area there is not a large amount of relevant research currently in existence, this is put down to the recent emergence of the technologies under discussion and it is assumed that more published sources will become available in the future.

4. How is AI Currently Regulated or Proposed to be Regulated?

Ultimately it is not the job of legislators to restrict and stymie technological breakthroughs, but it is their job to ensure these are done safely, ethically and responsibly (Jacobs, M. & Simon, J. 2023). In his article entitled *'AI Regulations Around the World: A Comprehensive Guide to Governing Artificial Intelligence'*, Savio Jacob highlights the different approaches currently being taken by different governments around the world. For example, the UK has produced a White Paper that proposes an innovation led approach to governance and regulation, whereas, US is currently taking a decentralised approach and has several sector specific agencies that set out to address the challenges brought on by AI development. In the EU, regulation is broken down in risk factors and very much geared to placing accountability for responsible development on the companies who develop these models first and the people who use the technology second. China is approaching this differently in the wish to remain at the forefront of AI technological developments, indeed their focus is on maintaining their superiority in AI and the plan under the Chinese Cybersecurity Law and the New Generation AI Development Plan integrated both security and data protection measures. Australia has produced the 'National Artificial Intelligence Ethics Framework', which direct the ethical principles by which AI technologies are developed and implemented, this approach focuses away from the data protection element for the individual user and places more emphasis on how the technology is developed (Jacob, 2024).

In March 2023 the Government of the UK published its White Paper entitled 'A Pro Innovation Approach to AI Regulation", indeed White Papers are often seen as a precursor to specific policy directions by which the government of the day wishes to pursue. The title of this specific paper championing the phrase 'Pro Innovation' is potentially problematic when it comes to developing a safety-first approach to regulation. Where innovation should certainly not be frustrated by regulation, striking a balance between what is innovative and what is ethical should always be considered, for example it could be argued that the approach CA took in 2016 to its campaigns was indeed innovative but was certainly not ethical. Therefore, we must throw a note of caution in pursuing an innovation led approach to regulating this industry, not simply allowing progression before safeguards are put in place to ensure everyone's safety and to this transparency is key.

In the White Paper the UK government recognises the importance of AI and the benefits already being seen in society, and it identified AI as being one of the five critical technologies. It recognises that across the world government are beginning to legislate for this and Sir Patrick Vallance (the then UK Governments Chief Scientific Advisor), notes that this is very much a time critical venture if the UK is to take full commercial advantage in a pro-innovation approach to make the UK an attractive place to encourage AI start-ups. The Paper emphasises the risks to both physical and mental health in addition to the need for data protection and privacy as well as the protection of human rights. Building public trust is also seen as a key factor in this White Paper, public trust is seen as a key to effective regulation with the need for clear and consistent support where government investment i.e. taxpayers' money is used. However, it is important to accelerate the adoption of AI across the UK, again putting a pro-innovation strategy at the forefront and promoting AI technologies in the most positive light possible. The danger in this approach that significant risks are either overlooked or simply dismissed. There is recognition that currently AI as a general-purpose technology is regulated across a complex framework of requirements in different sectors and industries and this is a concern and therefore the paper recognises the need for a clear and unified approach to regulation (UK Government, 2023).

In the US, the approach to governance is far more fragmented due to the nature of having both Federal and State Laws. It is entirely possible without blanket legislation for individual states to pass their own AI bills meaning that different states in the US will regulate in different ways. Currently in the state of California which is one of the earlier adopters of generative AI and with plans to help use it for things such as traffic management and highway monitoring, Senator Scott Wiener has authored a Bill proposing to create a stronger regulatory requirement, setting out further safety standards that would be designed to ensure models are thoroughly tested and to bring forward measures to prevent large AI models from being used to create potential catastrophes such as manipulation of the power grid or the building of chemical weapons (Nguyen, 2024).

On the 1st August 2024, the AI Act came into force across the EU. It classifies AI according to its assumed risk and identifies four categories of risk by which the Act seeks to impose regulation:

- **Unacceptable risk**
- **High risk systems**
- **Limited risk systems**
- **Minimal risk systems**

This Act pushes out further than just the boundaries of the EU by dictating that products intended to be put into service inside the EU regardless as to where they are based, are obliged to conform to the regulations laid down in the Act (Stewart, L. S. 2024). This also includes the products outputs, so even output generated from a system based outside of the EU, if that output is to be used within the EU it is covered by the Act (European Union, 2024).

2.1 Unacceptable Risk

Under the Act these are some examples of types of AI systems are prohibited i.e. not permitted and are covered by Article 5 of the Act. Unacceptable risk is likely to be the most challenging category to enforce unless the categorisation is very clearly defined, under EU law this would prohibit the development of systems that fall into this category and may bring legal challenges from any company whose systems are categorised in this way.

2.2 High Risk

Under Annex III of the AI Act these are some examples of areas that AI systems could be developed and deployed that are deemed high risk and therefore subject to significant regulation. The areas covered under the high-risk classification include some sectors under government control. Protecting critical national infrastructure such as transportation, power and water supply would be seen as a priority and therefore the introduction of AI systems into the running of any of these would need to be independently monitored. This would be especially true in terms of power generation where private investment and influence is substantial and where the need to generate profit is seen as a priority.

2.3 Penalties

Unlike so many Acts that are put in place to regulate and govern industries this one does have some teeth and is able to levy significant penalties where appropriate. For example, practices relating to Article 5 i.e., the prohibited Annex of the Act can lead to fines being levied in the region of €35,000,000 or, should the offender be proposing or promising something that is deemed prohibited 7% of previous annual worldwide turnover (whichever is higher). Where non-compliance is deemed according to elements of the Act not covered in Article 5 then fines of up to €15,000,000 or, if there is an undertaking then 3% of previous annual worldwide turnover (whichever is higher).

2.4 Codes of Practice and Governance

The AI Act does not introduce a Code of Practice but does make provision for one to be drawn up under article 95. It suggests that the AI office and member states collaborate to develop this as a voluntary application to governance of AI systems that are not deemed high risk (these systems are already covered under Chapter III of the AI Act). Article 95 also suggests that individual codes of practice may be drawn up by providers, deployers of AI systems or by interested stakeholders or representative organisations. This gives cause for concern as to how effective these could be, although it is recognised that producing a standardised one size fits all Code of Conduct may prove restrictive and prohibitive enough to encourage non-compliance or take away the incentive to sign up altogether.

2.5 ISO 42001

The official title of this standard is ISO/IEC 42001, Information Technology – Artificial Intelligence – Management System. The standard seeks to specify the requirements and to provide guidance to companies in relation to implementing AI systems for management purposes within their organisations and was published in 2023. This Standard does not set out best practice for the development of AI but more specifically how it is implemented and managed in the managements systems of an organisation. This is an important, although not fool proof, safeguard and with the growth in the use of AI by large companies and governments having an agreed standard to work to, will provide some semblance of oversight.

The Standard intends to help organisations to be responsible in their approach to the implementation of AI with specific considerations given to:

- The use of AI for decision making
- The use of AI for data analysis
- AI systems that perform continuous learning and change their behaviours during use

The Standard gives expectations to organisations to focus their effort into these areas to try to always ensure transparency in relation to how AI is being used and to attempt to mitigate the risks involved (Internation Standards Organisaton, 2023)

2.6 The 1st AI Safety Summit, November 2023

The AI Safety Summit was convened by the then Prime Minister of the UK Rishi Sunak in November 2023 at Bletchley Park, a site made famous after World War 2 as being the place where many Axis Codes were broken and certainly known to be one of the first uses of what we might recognise as modern computer systems. The Summit was attended by representatives from 29 different governments including the EU and the United States. The overarching aim of the Summit was to attempt to garner a joined-up approach in the future developments of AI, especially the development of “Frontier AI”.

2.7 The Bletchley Declaration

With Summits of this type, it is highly unlikely that we will get substantial and robust policy made, that all participants can agree on and where real change can happen. This Summit appears to have been no exception where a list of “nice to haves” was produced but not backed up by any enforcement, regulation or guidance to companies who are developing these advanced models. The Summit recognised, quite correctly the use of AI to enhance our lives whilst also recognised the need for safe design and how the next generation of frontier AI should be very ‘Human Centric’ (Government of the United Kingdom, 2023). It also recognises there are unseen risks. Referring to its concerns over current AI systems to manipulate or generate content that is inaccurate or deceptive in nature and notes the urgency of addressing these issues. What the Summit does not do in any substantial way is plot a route to any potential governance of this industry, indeed, governance and regulation is not mentioned until the third page of the document. When this is mentioned, it is vague, discussing applicable legal frameworks and national circumstances and seems to only pay lip service to the potential problem.

2.8 The 2nd AI Safety Summit, May 2024

The 2nd AI Safety Summit was held in Seoul, South Korea, was co-hosted by the Republic of Korea and the government of the UK. The outcome of the Summit was criticised by some in diluting the purpose of the Summit by adding additional topics for discussion other than simply AI safety, including innovation and inclusivity. Whilst this Summit could be seen as largely irrelevant, especially when compared to its previous outing it did produce several significant points of discussion. Most notably on AI safety institutes and their increasing number giving increased capacity across governments to better understand and regulate at both national and international levels. The addition of further AI safety institutes across countries such as South Korea, Japan and Canada when added to the already declared institutes in the UK and US, the European Commission AI Office serves as the AI safety institute for its 27 member countries, the Summit was able to secure the backing of 10 countries plus the EU for greater cooperation. This marks the true legacy of this Summit as to realising international cooperative efforts in relation to AI safety (Allen and Adamson, 2024).

2.9 AI companies and Their Responsibilities - Safety Committees

As noted in the EU AI Act responsibility is placed on the developers of this technology to develop systems that is not only safe but ethical. If we look at some of the most significant companies in the field, for example,

OpenAI in May 2024 introduced its Safety and Security Committee. This Committee is primarily responsible for reporting to the full board on critical safety issues and security discussion. This Committee is also tasked with reviewing the processes and safeguards around the development of new frontier AI models, reports from this Committee are set to be published in line with their recommendations to the board of OpenAI (OpenAI, 2024).

The Safety Committee for OpenAI, should ensure that the company does not go beyond what is deemed acceptable by law, however, there is a significant weakness in the makeup of the Committee. If we examine the members of the Committee it is comprised of existing board members which includes Sam Altman the CEO of OpenAI. Whilst it is not impossible for this Committee to do its job in a transparent manner, it must also be pointed out that there are conflicts of interest here and the lack of a truly independent overseer is significant.

3. Conclusions and Recommendations

Whilst researching this topic it became clear that that is a significant lack of a cooperative approach across the world to regulating anything, what is in one countries interest is not in another's, however, when it comes to technology such as AI, which has such potential, seeing this technology being exploited for any nefarious purpose is troubling. The question that has been running through this Paper is, of course, what we as a global community do about it. If we look at other areas of technology that have such potential for good, but also harm, and see how this is done and we may start to be able to develop answers to this problem. Atomic energy for example, has an international regulator, the International Atomic Energy Agency (IAEA) that has been on statute since 1956 with the United Nations and has a briefing to promote best practice, safety standards and to assist all economies in applying these standards. It would seem, a similar approach could be taken with AI development and therefore based upon this research it is recommended that:

- There is an implementation of an International Agency responsible for the application of agreed standards in AI development, this Agency needs to be properly funded through a similar model to the IAEA and have powers of inspection and sanction
- The Agency should have the power to launch a surprise inspection, providing it can be shown it has clear grounds to do so
- The Agency should seek to implement a Standardisation Programme that could be ISO42001, or based upon this, that AI developers should be able to sign up for and be regularly assessed to ensure that they are meeting these set standards
- Standards should be reviewed over a regular period, which is to be defined, to ensure that the Agency is working with the most up to date information and models

Whilst setting up an Agency of this type may prove significant, like all codes of practice it will require buy in from everyone to make it truly successful. However, the phrase ‘do not let the perfect be the enemy of the good’, is applicable if in doing this we can considerably reduce the risks posed by rogue AI development.

Other areas of the world that are likely to have significant influence on the development of AI include both the EU and the UK. However, the approach to AI regulation is significantly different between the two, the White Paper published by the UK seems to prioritise innovation above safety, which implies a potentially lighter regulatory touch. This approach is likely designed to encourage companies to locate the research and development arms in the UK as opposed to other nations thus ensuring significant economic benefit for the UK. Whereas the EU approach is seen as significantly more safety oriented, better defining both risks and penalties and the creation of an overseer that will have power at Federal level to impose significant penalties.

We have already seen the potential damage that even the less sophisticated AI type systems can do or at least have the potential to do in the CA scandal that came to light in 2016. Although the argument is that this was not true AI, the building blocks and potential can be easily observed in the use of algorithms designed to specially characterise individuals based on learned criteria. What was also apparent from this scandal is that although personal data is protected under frameworks such as GDPR this was not a significant deterrent and was exploited significant (Delcker, J. 2020).

Acknowledgements

This work was supervised and supported by Dr Richard Ward and Dr Mabrouka Abuhmida of the University of South Wales and Mrs Dawn Bradley of Coleg Gwent for proofreading.

References

Adamson, G. C. Allen, G. (2024). *The AI Seoul Summit*. Retrieved July 22, 2024, from https://www.csis.org/analysis/ai-seoul-summit

Buruiană, A. (2021) THE LEGAL CONSEQUENCES OF THE FACEBOOK-CAMBRIDGE ANALYTICA SCANDAL ON THE EU INFORMATION SOCIETY; THE PROPOSAL OF E-PRIVACY REGULATION. *Analele Universității Titu Maiorescu. Drept*. XX (XX), 47–58.

Delcker, J. (2020). *AI: Decoded: How Cambridge Analytica used AI*. Retrieved July 22, 2024, from https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-cambridge-analytica-used-ai-no-google-didnt-call-for-a-ban-on-face-recognition-restricting-ai exports/#:~:text=HOW%20CAMBRIDGE%20ANALYTICA%20USED%20AI%3A%20Artificial%20intelligence%20p

Chauhan, K. D. S. (2023) From 'What' and 'Why' to 'How': An Imperative Driven Approach to Mechanics of AI Regulation. *Global jurist*. [Online] 23 (2), 99–124.

European Union. (2024). *AI Act Explorer*. Retrieved July 16, 2024, from https://artificialintelligenceact.eu/ai-act-explorer/

Government of the United Kingdom. (2023). *About the AI Safety Summit 2023*. Retrieved May 29, 2024, from https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about

Ibrahim, Y. & Safieddine, F. (eds.) (2020) *Fake news in an era of social media : tracking viral contagion*. London ; Rowman & Littlefield International.

International Standards Organisaton. (2023). *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system*. Retrieved August 22, 2024, from https://www.iso.org/standard/81230.html

Ipsos. (2022). *IPSOS GLOBAL TRUSTWORTHINESS MONITOR, STABILITY IN AN UNSTABLE WORLD.* Ipsos.

Jacob, S. (2024). *AI Regulations Around the World: A Comprehensive Guide To Governing Artificial Intelligence*. Retrieved July 10, 2024, from https://www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-the-world/

Jaquemard, T. (2023). *Frontier AI: Heading safely into new territory*. Retrieved May 29, 2024, from https://trilateralresearch.com/emerging-technology/frontier-ai-heading-safely-into-new-territory

Jacobs, M. & Simon, J. (2023) Reexamining computer ethics in light of AI systems and AI regulation. *Ai and ethics (Online)*. [Online] 3 (4), 1203–1213.

Matara, M. et al. (2022) *Facebook and Cambridge Analytica : an unholy alliance*. London: The Eugene D. Fanning Center for Business Communication, Mendoza College of Business, University of Notre Dame.

Nguyen, T. (2024). *California advances unique safety regulations for AI companies despite tech firm opposition*. Retrieved August 5, 2024, from https://apnews.com/article/california-meta-google-ai-regulation-fight-e3a9c25b597b5f469fe0de149572f87e

Nyabola, N. (2023) Seeing the forest – and the trees: The global challenge of regulating social media for democracy. *The South African journal of international affairs*. [Online] 30 (3), 455–471.

Open AI. (2024). *OpenAI Board Forms Safety and Security Committee*. Retrieved July 22, 2024, from https://openai.com/index/openai-board-forms-safety-and-security-committee/

Pinnell, J. (2024) AI, DISINFORMATION AND DEMOCRACY: THE NEED FOR PARLIAMENTS TO ACT. *Parliamentarian*. 105 (1), 36-.

Schneble, C. O. et al. (2018) The Cambridge Analytica affair and Internet-mediated research. *EMBO reports*. [Online] 19 (8), .

Stewart, L. S. (2024) The regulation of AI-based migration technologies under the EU AI Act: (Still) operating in the shadows? *European law journal : review of European law in context*. [Online] 30 (1/2), 122–135.

UK Government. (2023). *A pro-innovation approach to AI regulation*. Retrieved July 22, 2024, from https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#executive-summary