

# **When Stochastic Parrots Learn to Swim: The Regulation of General Purpose Artificial Intelligence in the EU**

Vartul Srivastava

Master's Thesis in European and International Trade Law

HARN63

Spring 2023



**SCHOOL OF  
ECONOMICS AND  
MANAGEMENT**

# Table of Contents

<b>1. Introduction.....</b>	<b>4</b>
1.1 Background .....	4
1.2 Purpose and research question .....	8
1.3 Delimitations .....	8
1.4 Materials and method .....	9
1.5 Structure .....	9
<b>2. The Many and Complex Dangers of GPAIs.....</b>	<b>11</b>
<b>3. The EU’s strategy to adapt the AI Act to GPAIs.....</b>	<b>18</b>
3.1 GPAI Provisions in the AI Act.....	18
3.1.1 Open Source GPAIs – A complicated problem .....	23
3.2 GPAI Provisions in the AI Act EP Version .....	25
<b>4. Remedies and Suggestions to Address the Critical Failings of the AI Act</b>	
<b>30</b>	
4.1 Would the AI Act benefit from a regulatory focus on users? .....	30
4.2 Addressing the Content Problem of GPAIs .....	33
<b>5. Exploring Alternative Regulatory Approaches .....</b>	<b>35</b>
5.1 Is a technology-neutral law the way forward? .....	35
5.2 The Choice between Legal Certainty and Regulatory Flexibility .....	37
5.3 Regulating Stochastic Parrots that have Learned to Swim .....	40
5.4 From a Purpose-Driven Paradigm of Regulation to a Purpose-Driven Paradigm of Regulation .....	45
<b>6. Conclusion .....</b>	<b>49</b>
<b>Reference list / Bibliography.....</b>	<b>51</b>

## **Abstract**

General Purpose AI systems like ChatGPT pose unique dangers to society owing to the rapid evolution in their capabilities, and their ability to be utilized for a large variety of purposes, including those unintended by the developers of these AI systems. This thesis examines those dangers, and critically analyses the EU's draft AI Act and its amendments with a view to ascertain if the draft law adequately tackles those dangers. It finds that the draft AI Act is ill-equipped to effectively regulate general purpose AI systems, and suffers from issues of rigidity, onerous regulatory obligations, and a lack of consideration of the unique characteristics of general purpose AI systems. The thesis explores and critiques the solutions offered by scholars to remedy these issues, and finally proposes an alternative regulatory approach focused on a goal-based regulation that would empower the European Commission to frame and amend rules governing general purpose AI systems, as well as the AI space at large.

## Abbreviations

EU	European Union
Commission	European Commission
Council	Council of the European Union
AI Act	The Artificial Intelligence Act
AI Act EP Version	The Artificial Intelligence Act European Parliament Version
AI	Artificial Intelligence
GPAI	General Purpose Artificial Intelligence
Parliament	The European Parliament
LM	Language Model
SMEs	Small and Medium Enterprises

# 1. Introduction

## 1.1 Background

In April 2021, the European Commission (Commission) published a draft for an AI (Artificial Intelligence) Act, with the purpose being to regulate artificial intelligence within the European Union.<sup>1</sup> This was the outcome of years of discussions and whitepapers on the subject, and the beginning of the process to formalize rules to govern AI.<sup>2</sup> The draft defined what constituted an AI system<sup>3</sup>, and had a classification system that categorized AI according to the risk it posed to society.<sup>4</sup> What followed was months of work by the EU decision-making institutions (namely, the Commission, the EU Parliament, and the Council of the EU) to negotiate and draft the core provisions of the Act.<sup>5</sup>

In November 2021, the Council shared the first compromise text on the AI Act draft with major changes<sup>6</sup>, which was followed by further amendments to the draft all through 2022.<sup>7</sup> Progress was measured and the adoption of finalized rules on AI was in sight.

In November 2022, OpenAI, an American artificial intelligence research and deployment company<sup>8</sup>, released ChatGPT, a conversational AI model (essentially

---

<sup>1</sup> Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts', COM (2021) 206 final.

<sup>2</sup> European commission, 'A European approach to artificial intelligence' (Shaping Europe's digital future, 26 January 2023) <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 25 May 2023.

<sup>3</sup> Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts', COM (2021) 206 final, page 39.

<sup>4</sup> Ibid, page 45.

<sup>5</sup> Future of life institute, 'Developments' (The Artificial Intelligence Act, N/A) <https://artificialintelligenceact.eu/developments/> accessed 25 May 2023.

<sup>6</sup> Bertuzzi L, 'EU Council Presidency Pitches Significant Changes to AI Act Proposal' (www.euractiv.com, 1 December 2021) <https://www.euractiv.com/section/digital/news/eu-council-presidency-pitches-significant-changes-to-ai-act-proposal/> accessed 25 May 2023.

<sup>7</sup> Future of life institute, 'Developments' (The Artificial Intelligence Act, N/A) <https://artificialintelligenceact.eu/developments/> accessed 25 May 2023.

<sup>8</sup> OpenAI, 'About' (OpenAI) <https://openai.com/about> accessed 25 May 2023.

a chatbot), to the public. Within two months of launch, the app reached a 100 million users, and became the fastest growing app in history.<sup>9</sup>

The launch of ChatGPT was nothing short of catastrophic to the EU's then-existing plans to regulate AI.<sup>10</sup> To understand why, it is imperative to discuss the original EU proposal for the regulatory framework on AI<sup>11</sup> (before ChatGPT forced major changes to it), and how ChatGPT is different from the chatbots that have existed since the 1960s<sup>12</sup>.

The foundational elements of the EU proposal emanate from a single statement on the commission's website –

*“While most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that we must address to avoid undesirable outcomes.”<sup>13</sup>*

This statement is so consequential that the entire EU proposal follows from it as a logical corollary. Assuming the statement to be correct, and following basic reasoning, the first task of a regulatory framework would likely be to categorize AI systems based on the kind of risk they pose. And this is exactly the approach the actual proposal takes – it defines four levels of risk: Unacceptable risk, High risk, Limited risk, and Minimal/no risk.<sup>14</sup> To take a few examples, Unacceptable risk in the Commission's definition includes AI systems that allow social scoring by governments, while High risk involves AI systems used in critical infrastructure

---

<sup>9</sup> Hu K, 'CHATGPT Sets Record for Fastest-Growing User Base - Analyst Note' (Reuters, 2 February 2023) <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> accessed 25 May 2023.

<sup>10</sup> Volpicelli G, 'CHATGPT Broke the EU Plan to Regulate AI' (POLITICO, 6 March 2023) <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/> accessed 25 May 2023.

<sup>11</sup> European Commission, 'Regulatory Framework Proposal on Artificial Intelligence' (Shaping Europe's digital future, 29 September 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 25 May 2023.

<sup>12</sup> ZEMČÍK MgrT, 'A Brief History of Chatbots' [2019] DEStech Transactions on Computer Science and Engineering 14 [https://www.researchgate.net/publication/336734161\\_A\\_Brief\\_History\\_of\\_Chatbots](https://www.researchgate.net/publication/336734161_A_Brief_History_of_Chatbots) accessed 25 May 2023, page 15.

<sup>13</sup> European Commission, 'Regulatory Framework Proposal on Artificial Intelligence' (Shaping Europe's digital future, 29 September 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 25 May 2023.

<sup>14</sup> Ibid.

like transport, safety concerns of products etc. Limited risk includes chatbots, while Minimal/no risk cites AI-enabled video games as an example.<sup>15</sup>

Having created this risk-based classification, the next logical step would be to impose specific obligations on AI systems depending on the classification they fall under. Naturally, AI systems that pose unacceptable risk would be disallowed, while, High risk systems would invite strict obligations, and so on. Again, this is exactly the route the proposal follows - AI systems that pose unacceptable risk would be banned, while High risk AI systems would be subjected to pre-release requirements regarding risk assessment, minimizing discriminatory outcomes, human oversight measures etc.<sup>16</sup> Limited risk AI systems would be given specific transparency obligations, and Minimal risk AI is allowed to be freely used.<sup>17</sup>

There's also a section in the proposal regarding making the legislation future-proof, but it's a broad statement lacking any details.<sup>18</sup>

ChatGPT is a chatbot. As such, on a plain reading of the proposal, it would be classified as Limited risk AI, and thus would only be required to follow certain transparency obligations (so users are aware that they are conversing with a machine and not another human being).<sup>19</sup>

Why that would be problematic can be demonstrated by the real-world use of ChatGPT. On 14<sup>th</sup> March 2023, OpenAI released ChatGPT-4 to the public – the latest model of their AI chatbot. Some of the developments that took place within weeks of the release of ChatGPT-4 are as follows:

1. Khan Academy, an American non-profit educational organization, demoed the use of GPT-4 to transform learning. Examples included a debate with the chatbot on whether the government should cancel student debt, chatting with a historical figure (like Isaac Newton) or a literary character (like the Greek god Zeus), and asking the chatbot's advice on navigating

---

<sup>15</sup> European Commission, 'Regulatory Framework Proposal on Artificial Intelligence' (Shaping Europe's digital future, 29 September 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 25 May 2023.

<sup>16</sup> Ibid.

<sup>17</sup> Ibid.

<sup>18</sup> Ibid.

<sup>19</sup> Ibid.

college admissions.<sup>20</sup> In the opinion of the author of this thesis, the chatbot's answers in these examples could be qualified as human-level, with no obvious errors.

2. Aaron Mulgrew, a security researcher at Forcepoint (a cybersecurity company), was able to get ChatGPT to generate a zero-day malware that would pretend to be a screensaver, and once downloaded on a user's system, would run itself and send documents and images from the user's drive to the attacker. It took Aaron a few hours to generate the malware, when it would've taken a team of malware developers a few weeks to build something similar.<sup>21</sup>
3. Be My Eyes, a mobile app designed to make the world more accessible for people with low-vision or blindness<sup>22</sup>, announced a virtual volunteer tool powered by ChatGPT-4. Users will be able to send images to the AI, which will then answer questions and provide assistance on a plethora of tasks.<sup>23</sup> Examples included guiding a blind person through a gym, describing the colours and features of clothes, identifying a plant, and more.<sup>24</sup>

The three use-cases described above are merely a tiny sample of the ways ChatGPT is being used today.

The original EU proposal to regulate AI was based on a purpose-driven understanding of AI. Every AI system was seen as having a purpose (or a set of purposes), and thus AI systems could be classified and regulated depending on the kinds of risks and dangers associated with the purpose they intended to serve.

ChatGPT, as evident from the above examples, can be used in many disparate and surprising ways. It does not have a singular purpose. As such, it belongs to category of AI systems known as General-Purpose AIs (GPAIs).

---

<sup>20</sup> Khan Academy, 'GPT-4 Khan Academy in Depth Demo' (YouTube, 14 March 2023) <https://www.youtube.com/watch?v=rnIgnS8Sug> accessed 25 May 2023.

<sup>21</sup> Mulgrew A, 'I Built a Zero Day Virus with Undetectable Exfiltration Using Only CHATGPT Prompts' (Forcepoint, 9 May 2023) <https://www.forcepoint.com/blog/x-labs/zero-day-exfiltration-using-chatgpt-prompts> accessed 25 May 2023.

<sup>22</sup> be my eyes, 'The Story about Be My Eyes' (The story about Be My Eyes) <https://www.bemyeyes.com/about> accessed 25 May 2023.

<sup>23</sup> be my eyes, 'Introducing Our Virtual Volunteer Tool for People Who Are Blind or Have Low Vision, Powered by OpenAI's GPT-4' (Be My Eyes - See the world together) <https://www.bemyeyes.com/blog/introducing-be-my-eyes-virtual-volunteer> accessed 25 May 2023.

<sup>24</sup> Ibid.



The purpose ChatGPT is used for can range from seemingly as simple and harmless as writing text for birthday invitations, to as complex and important as helping people with disabilities navigate the world, and to as dangerous as writing code for a malware or virus.

Thus the plurality of purposes that define General-Purpose AIs frustrate the purpose-based classification contained within the Commission's proposal.

The EU has decided to review GPAI systems like ChatGPT<sup>25</sup>, and is making changes to its draft AI Act to adapt to the challenges posed by GPAIs.<sup>26</sup>

In light of that, it would be instructive to discuss the specific risks posed by GPAIs to society, followed by an analysis of the latest publicly available drafts of the AI Act that purport to deal with those risks, as well as recent developments reported in the news regarding updates to the EU's plans regarding regulation of GPAIs.

## **1.2 Purpose and research question**

The purpose of this thesis is to analyse and critique the regulation of General Purpose AI (GPAI) systems in the EU's AI Act in context of the particular dangers posed by GPAI systems, and to explore if other regulatory choices would have better served the regulation of GPAI in the EU.

### **Research Questions:**

1. What are the dangers posed by GPAI systems?
2. How does the EU plan to regulate GPAI systems?
3. Could other regulatory choices have better governed GPAI systems?

## **1.3 Delimitations**

General Purpose AI (GPAI) systems raise important questions in a variety of disparate fields. The thesis does not analyse three major fields impacted by GPAI – Environment and Sustainability, Data Protection, and Intellectual Property. The principal reason for this is that the EU's AI Act, the focus of this thesis, does not

---

<sup>25</sup> Bayer M, 'EU Task Force to Review Chatgpt' (CIO, 24 April 2023) <https://www.cio.com/article/474610/eu-task-force-to-review-chatgpt.html> accessed 25 May 2023.

<sup>26</sup> Sharma S, 'EU Closes in on Ai Act with Last-Minute CHATGPT-Related Adjustments' (Computerworld, 28 April 2023) <https://www.computerworld.com/article/3695009/eu-closes-in-on-ai-act-with-last-minute-chatgpt-related-adjustments.html> accessed 25 May 2023.

regulate the interactions of GPAI systems with any of those fields. Further, the interactions of GPAI with Intellectual Property Law are complex and expansive, and deserve a dedicated paper to enable a detailed analysis.

## **1.4 Materials and method**

The thesis follows the legal dogmatic method. In the examination of the primary sources, the following drafts and amendments to the proposed Artificial Intelligence Act in the European Union form the focus of the analysis-

- Commission, ‘Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts’, COM (2021) 206 final. (referred in this thesis as the AI Act)
- DRAFT Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council, Brando Benifei & Ioan-Dragoş Tudorache (May 9, 2023) (referred in this thesis as the AI Act EP Version)

In addition, the thesis conducts a critical examination of scholarly articles and reports to complement the analysis of the drafts of the AI Act and reach certain conclusions.

## **1.5 Structure**

The thesis is divided into six chapters. The first chapter is an introduction to the EU’s initial approach to regulating AI, and the complexities of GPAI systems which rendered that approach ineffective. The second chapter explores the dangers posed by GPAI systems that require regulation. The third chapter analyses and critiques the changes made by the EU to its draft AI Act to target GPAI systems. The fourth chapter suggests remedies for the issues that plague the draft AI Act in its attempt to regulate GPAIs. The fifth chapter explores alternative regulatory approaches, and proposes a goal-based regulatory approach that empowers a regulatory agency and emphasises flexibility. It also examines the risks associated with such an approach. The final chapter is the conclusion, which summarizes the findings of the thesis.



## 2. The Many and Complex Dangers of GPAIs

In early 2021, the lead researchers of Google’s AI ethics team published a paper titled ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?’<sup>27</sup>. The ‘language models’ referenced in the title are the technology GPAIs are built upon.<sup>28</sup> In the paper, the researchers highlight the dangers of assigning comprehension and meaning to the outputs of language models (LMs)-

*“Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot.”<sup>29</sup>*

The authors highlight how the training data derived from the internet results in encoding of hegemonic views within the language model, and the tendency of these language models to then amplify biases that were present in the data they were trained upon.<sup>30</sup> They then explore the risks of releasing such a language model into the world at scale, and state that *“the mix of human biases and seemingly coherent language heightens the potential for automation bias, deliberate misuse, and amplification of a hegemonic worldview.”<sup>31</sup>*

The release of this paper resulted in the firing of its authors from Google when they refused to retract the paper.<sup>32</sup>

---

<sup>27</sup> Bender EM and others, ‘On the Dangers of Stochastic Parrots’ [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610 <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 25 May 2023.

<sup>28</sup> Muehmel K, ‘What Is a Large Language Model, the Tech behind Chatgpt?’ (Blog, 15 March 2023) <https://blog.dataiku.com/large-language-model-chatgpt> accessed 25 May 2023.

<sup>29</sup> Bender EM and others, ‘On the Dangers of Stochastic Parrots’ [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610 <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 25 May 2023, pages 616 and 617.

<sup>30</sup> Ibid, page 616.

<sup>31</sup> Ibid.

<sup>32</sup> ‘Margaret Mitchell: Google Fires Ai Ethics Founder’ (BBC News, 20 February 2021) <https://www.bbc.com/news/technology-56135817> accessed 25 May 2023.

GPAIs may indeed be merely stochastic parrots, creating content without comprehension, but as shall be discussed in this chapter, that only amplifies the dangers associated with them.

The Future of Life Institute (an independent non-profit working to highlight both the potential and risks of AI<sup>33</sup>) defines GPAIs succinctly as follows-

*“General purpose AI systems are AI systems that have a wide range of possible uses, both intended and unintended by the developers. They can be applied to many different tasks in various fields, often without substantial modification and fine-tuning.”*<sup>34</sup>

The thesis considers this definition to include, in addition to models like ChatGPT, AI systems that generate images (like Midjourney, Dall-E and Stable Diffusion) and videos (like Elai.io, Pictory and Synthesia), since like text, images and videos are simply mediums to convey information, and may be used in a wide variety of scenarios for various purposes.

Before delving further into the risks and challenges associated with GPAIs, it’s important to emphasize the potentially transformative benefits these systems are capable of bringing to the world. GPAIs could be used to aid climate research<sup>35</sup>, in healthcare and medicine<sup>36</sup>, in urban planning<sup>37</sup>, in robotics<sup>38</sup>, and have the potential to revolutionize education<sup>39</sup> amongst many other existing and future uses.

---

<sup>33</sup> Future of life institute, 'About' (The Artificial Intelligence Act, N/A) <https://artificialintelligenceact.eu/about/> accessed 25 May 2023.

<sup>34</sup> Future of Life Institute, 'General Purpose AI and the AI Act' (The Artificial Intelligence Act, May 2022) <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf> accessed 25 May 2023.

<sup>35</sup> Biswas SS, 'Potential Use of Chat GPT in Global Warming - Annals of Biomedical Engineering' (SpringerLink, 1 March 2023) <https://link.springer.com/article/10.1007/s10439-023-03171-8> accessed 25 May 2023.

<sup>36</sup> Sallam M, 'ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns' (2023) 11 Healthcare 887 <http://dx.doi.org/10.3390/healthcare11060887> accessed 25 May 2023.

<sup>37</sup> Wang D, Lu C-T and Fu Y, 'Towards Automated Urban Planning: When Generative and CHATGPT-like AI Meets Urban Planning' (arXiv.org, 8 April 2023) <https://arxiv.org/abs/2304.03892> accessed 25 May 2023.

<sup>38</sup> Vemprala S and others, 'ChatGPT for Robotics: Design Principles and Model Abilities' (Microsoft, 20 February 2023) [https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT\\_Robotics.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT_Robotics.pdf) accessed 25 May 2023.

<sup>39</sup> Adiguzel T, Kaya MH and Cansu FK, 'Revolutionizing Education with AI: Exploring the Transformative Potential of CHATGPT' (2023) 15 Contemporary Educational Technology <https://www.cedtech.net/download/revolutionizing-education-with-ai-exploring-the-transformative-potential-of-chatgpt-13152.pdf> accessed 25 May 2023.

As such, the exploration of the dangers of GPAIs takes place in the context of these benefits.

One useful way to explore some of the major risks would be to deconstruct and analyse the aforementioned definition of GPAIs:

The first statement in the definition mentions that GPAI systems have a wide range of possible uses, which may not necessarily be intended by the developers.

This is already being seen in the case of ChatGPT. In the hands of users, it has been made to write and debug programming code, translate text from one language to the other, solve math questions, compose music, and even play tic-tac-toe.<sup>40</sup> It has been used by users as a therapist<sup>41</sup>, as a tool to design arcade games<sup>42</sup>, to build an online business<sup>43</sup>, and ironically, to solve captchas used by certain websites to verify that the user is a human being and not a bot<sup>44</sup>.

Since the capabilities of GPAIs make it feasible to use them for malicious purposes (like building malware or ransomware, generating harmful or biased content like scam emails etc.<sup>45</sup>), it is logical to place an onus on developers to restrict such malicious uses of their product. In the case of ChatGPT, OpenAI has integrated guardrails into the system that should prevent it from generating harmful, biased, abusive or dangerous content.<sup>46</sup> However, users have managed to get around those guardrails to generate a wide variety of problematic content. Early in 2023, hackers devised a method to bypass those guardrails so they could make ChatGPT generate

---

<sup>40</sup> Gangwar A, '12 Cool Things You Can Do with Chatgpt' (Beebom, 29 March 2023) <https://beebom.com/cool-things-chatgpt/> accessed 25 May 2023.

<sup>41</sup> Pirnay E, 'We Spoke to People Who Started Using CHATGPT as Their Therapist' (VICE, 27 April 2023) <https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist> accessed 25 May 2023.

<sup>42</sup> Scott G, 'Last Night I Made a Website That Uses GPT-4 to Code Any Arcade Game You Can Think of and Let You Play It Instantly. Here's a Demo of the Infinite Arcade. If People like It, I'll Publish the Site Later Today or Tomorrow!' (Twitter, 16 March 2023) <https://twitter.com/thegarrettscott/status/1636477569565335553> accessed 25 May 2023.

<sup>43</sup> Fall JG, 'How It Started:•3,713 on Twitter•0 on Discord•\$0 Revenue How It's Going:•87,903 on Twitter•1,501 on Discord•\$130 Revenue' (Twitter, 19 March 2023) <https://twitter.com/jacksonfall/status/1637459175512092672> accessed 25 May 2023.

<sup>44</sup> Abraham TM, 'I Got to Try GPT-4's Multimodal Capabilities and It's Quite Impressive! A Quick Thread of Examples...Let's Start out with Solving a CAPTCHA, No Big Deal' (Twitter, 16 March 2023) <https://twitter.com/iScienceLuvr/status/1636479850214232064> accessed 25 May 2023.

<sup>45</sup> White MJ, 'The Dark Side of CHATGPT: Things It Can Do, but It Shouldn't' (Digital Trends, 16 April 2023) <https://www.digitaltrends.com/computing/bad-things-chatgpt-has-been-used-for/> accessed 25 May 2023.

<sup>46</sup> OpenAI (Our approach to AI safety, 5 April 2023) <https://openai.com/blog/our-approach-to-ai-safety> accessed 25 May 2023.

everything from phishing emails to malware and viruses, and then began to sell that service on an underground crime forum.<sup>47</sup>

There exist online a plethora of websites dedicated to helping users easily bypass ChatGPT's content restrictions<sup>48 49</sup>, and OpenAI seems to be in an arms race to shut down these bypass measures even as new ones appear every day.<sup>50</sup>

And the issue isn't limited to text-based GPAIs. Image-generating AIs like Stable Diffusion do officially come with filters to block certain kinds of content (for example, NSFW - Not Safe for Work content), but these restrictions can easily be circumvented due to the open-source nature of the code<sup>51</sup>, and there are plenty of videos online showing how to do so<sup>52</sup>.

The second statement in the definition of GPAI states that these systems can be used in a variety of fields to perform a multitude of tasks, usually without the requirement of any major modifications to the system.

This is clearly borne out in the above examples. It may be argued that many of the risky uses of GPAIs, whether they be building computer viruses, generating fake news, or otherwise problematic content, could be achieved in the past without AI as well. And while that is true, it is the ease with which GPAIs can be used for a number of malicious purposes that is the crux of the problem. A tool at one's fingertips that automates either a large part of a complex process or the entirety of it is far more likely to result in the widespread generation of problematic content

---

<sup>47</sup> Goodin D, 'Hackers Are Selling a Service That Bypasses CHATGPT Restrictions on Malware' (Ars Technica, 8 February 2023) <https://arstechnica.com/information-technology/2023/02/now-open-fee-based-telegram-service-that-uses-chatgpt-to-generate-malware/> accessed 25 May 2023.

<sup>48</sup> Ankita, 'How to Bypass CHATGPT Filter Restrictions' (MLYearning, 9 May 2023) <https://www.mlyearning.org/how-bypass-chat-gpt-restrictions/> accessed 25 May 2023.

<sup>49</sup> Goodin D, 'Hackers Are Selling a Service That Bypasses CHATGPT Restrictions on Malware' (Ars Technica, 8 February 2023) <https://dataconomy.com/2023/05/04/how-to-remove-chatgpt-restrictions/> accessed 25 May 2023.

<sup>50</sup> OpenAI (Our approach to AI safety, 5 April 2023) <https://openai.com/blog/our-approach-to-ai-safety> accessed 25 May 2023.

<sup>51</sup> Edwards B, 'With Stable Diffusion, You May Never Believe What You See Online Again' (Ars Technica, 6 September 2022) <https://arstechnica.com/information-technology/2022/09/with-stable-diffusion-you-may-never-believe-what-you-see-online-again/3/> accessed 25 May 2023.

<sup>52</sup> Goud B, 'Game Changing AI - It's Completely Free!' (YouTube, 6 March 2023) <https://www.youtube.com/watch?v=HUB2tFI5rqM&t=572s> accessed 25 May 2023.

than if the content could only be generated manually and painstakingly by certain domain-experts.<sup>53</sup>

This ease of application has resulted in GPAIs now being used to generate fake news<sup>54</sup> and fake videos<sup>55</sup> in prominent election campaigns across the world, and grim predictions have been made regarding the consequences of the emergence of AI on democratic systems.<sup>56</sup>

However, malicious uses of GPAIs aren't the only ones that must invite considerations of strict regulation.

### Threat to livelihood-

There have been legitimate concerns that AI may end up replacing a variety of workers as it swallows up a large number of jobs previously performed by human beings – be it ophthalmologists<sup>57</sup>, journalists<sup>58</sup>, teachers<sup>59</sup>, and swaths of other workers engaged in routine jobs, leading to unemployment and a rise in income inequality<sup>60</sup>. A 2023 study by a researcher at Indiana University ended up concluding that AI systems like ChatGPT are likely to have a full impact on 32.8% of occupations, and a partial impact on 36.5% of occupations.<sup>61</sup>

Returning to the original EU proposal for a regulatory framework on AI, it is worth noting that it designates AI systems that are considered to be “*a clear threat to the*

---

<sup>53</sup> Brundage M and others, ‘The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation’ (arXiv.org, 20 February 2018) <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf> accessed 25 May 2023, page 5.

<sup>54</sup> Khan S, ‘What’s Fake and Not Fake in Turkey’s Election’ (Inkstick, 16 May 2023) <https://inkstickmedia.com/whats-fake-and-not-fake-in-turkeys-election/> accessed 25 May 2023.

<sup>55</sup> Elci A, ‘Ai Deepfakes and Disinformation Loom over Turkey’s Election’ (euronews, 12 May 2023) <https://www.euronews.com/next/2023/05/12/ai-content-deepfakes-meddling-in-turkey-elections-experts-warn-its-just-the-beginning> accessed 25 May 2023.

<sup>56</sup> Baecker C and others, ‘Threats Provided by Artificial Intelligence That Could Disrupt the Democratic System’ (OPUS 4 | Threats provided by artificial intelligence that could disrupt the democratic system, 1 January 1970) <https://opus4.kobv.de/opus4-fhbrb/frontdoor/index/index/docId/2898> accessed 25 May 2023.

<sup>57</sup> Korot E and others, ‘Will AI Replace Ophthalmologists?’ (Translational Vision Science & Technology, 28 January 2020) <https://tvst.arvojournals.org/article.aspx?articleid=2760332> accessed 25 May 2023.

<sup>58</sup> Miroshnichenko A, ‘Ai to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is “Yes”)’ (MDPI, 23 July 2018) <https://www.mdpi.com/2078-2489/9/7/183> accessed 25 May 2023.

<sup>59</sup> Alam A, ‘Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education’ [2021] 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3) <https://ieeexplore.ieee.org/abstract/document/9697300> accessed 25 May 2023.

<sup>60</sup> Zarifonharvar A, ‘Economics of Chatgpt: A Labor Market View on the Occupational Impact of Artificial Intelligence’ [2023] SSRN Electronic Journal [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4350925](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350925) accessed 25 May 2023, page 26.

<sup>61</sup> Ibid.



*safety, livelihoods and rights of people*”<sup>62</sup> as systems that pose unacceptable risks, and will thus be banned.

#### Tendency to generate plausible yet false information-

GPAIs suffer from a particular flaw known as hallucination. They sometimes generate responses that appear confident and plausible, but have no basis in truth.<sup>63</sup> The issue stems from the very structure of these algorithms – they are designed to predict the most probable word that should come next in a sequence<sup>64</sup>, and, according to some AI experts, have no sense of truth<sup>65</sup>.

The tendency to generate false but plausible responses can be particularly dangerous as more users come to rely on these systems, especially in sensitive areas like the legal field or healthcare.<sup>66</sup>

#### Tendency to generate discriminatory responses-

GPAIs also have been shown to produce responses or content that may be biased and discriminatory in a variety of ways.

GPAIs are trained on a huge amount of data, which contains historical and societal biases, and this can affect the output of these systems.<sup>67</sup> ChatGPT, for example, was trained on a massive dataset of text from the internet (websites, articles, books etc.) totalling 300 billion words.<sup>68</sup> The training data could contain all kinds of harmful and discriminatory content, which the AI might then internalize, and go on to generate outputs that reflect that problematic training data. In 2022, Lensa AI, an app that allows users to upload photos of anyone and then generates the likeness of

---

<sup>62</sup> European Commission, ‘Regulatory Framework Proposal on Artificial Intelligence’ (Shaping Europe’s digital future, 29 September 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 25 May 2023.

<sup>63</sup> Li Z, ‘The Dark Side of CHATGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination’ (arXiv.org, 21 April 2023) <https://arxiv.org/abs/2304.14347> accessed 25 May 2023, page 1.

<sup>64</sup> Ibid.

<sup>65</sup> Volpicelli G, ‘CHATGPT Broke the EU Plan to Regulate AI’ (POLITICO, 6 March 2023) <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/> accessed 25 May 2023.

<sup>66</sup> Li Z, ‘The Dark Side of CHATGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination’ (arXiv.org, 21 April 2023) <https://arxiv.org/abs/2304.14347> accessed 25 May 2023, page 1.

<sup>67</sup> Borji A, ‘A Categorical Archive of CHATGPT Failures’ (arXiv.org, 3 April 2023) <https://arxiv.org/abs/2302.03494> accessed 25 May 2023, pages 11-14.

<sup>68</sup> GPT Blogs, ‘CHATGPT: How Much Data Is Used in the Training Process?’ (GPT Blogs, 9 February 2023) <https://gptblogs.com/chatgpt-how-much-data-is-used-in-the-training-process> accessed 25 May 2023

that person in various forms of clothing, environments and situations, was criticised<sup>69</sup> for its generated female images, which often showed sexual poses or nudity, especially with photos of Asian women, while the men were shown as doctors and astronauts.<sup>70</sup>

To use an example, a GPAI trained on such biased data, when used by an employer to analyse the Curriculum Vitae of candidates for a position, might be less likely to recommend candidates from marginalized groups for the position, and might unduly favour candidates from historically advantaged groups.<sup>71</sup>

Bias would also be a serious issue when these AI systems are used in healthcare<sup>72</sup> and other sensitive fields.

It is to be noted that this is an illustrative, not an exhaustive list of the dangers posed by GPAI systems.

---

<sup>69</sup> Sparrentak KV, ‘Parliamentary Question: Lensa Image-Generation App: Concerns Regarding Bias, Harassment, Privacy, and EU and National Security: E-004128/2022: European Parliament’ (European Parliament, 21 December 2022) [https://www.europarl.europa.eu/doceo/document/E-9-2022-004128\\_EN.html](https://www.europarl.europa.eu/doceo/document/E-9-2022-004128_EN.html) accessed 25 May 2023.

<sup>70</sup> Heikkilä, M. (2022) The viral ai avatar app lensa undressed me-without my consent, MIT Technology Review. Available at: <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent> <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent> accessed 25 May 2023.

<sup>71</sup> Zhuo TY and others, ‘Exploring Ai Ethics of CHATGPT: A Diagnostic Analysis’ (arXiv.org, 22 February 2023) <https://arxiv.org/abs/2301.12867> accessed 25 May 2023, page 3.

<sup>72</sup> The Lancet Digital Health, ‘Chatgpt: Friend or Foe?’ (2023) 5 The Lancet Digital Health [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(23\)00023-7/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00023-7/fulltext) accessed 25 May 2023

### 3. The EU's strategy to adapt the AI Act to GPAIs

The latest drafts of the AI Act retain the focus on High-risk AI systems, and establish strict obligations for those systems, but also include specific provisions to regulate GPAIs. In the critique of those provisions, the thesis relies on the draft text adopted by the EU Council on Dec 6, 2022<sup>73</sup> (henceforth referred to as the AI Act) and the draft compromise amendments of May 9<sup>th</sup> 2023<sup>74</sup> negotiated in the European Parliament (henceforth referred to as AI Act EP version).

#### 3.1 GPAI Provisions in the AI Act

Recital 12c of the AI Act mentions general purpose AI systems and the plurality of contexts they are applicable in. It outlines the need to have specific requirements and obligations for such systems.

Article 3(1b) of the Act defines a general purpose AI system as –

*“an AI system that - irrespective of how it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems.”*

The definition has immediate issues. It states that a GPAI is an AI system intended to perform functions such as image and speech recognition. The definition goes on to mention other capabilities, but notably the capabilities do not appear to be

---

<sup>73</sup> General approach adopted by the EU Council on Dec. 6, 2022 <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf> accessed 25 May 2023.

<sup>74</sup> DRAFT Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council, Brando Benifei & Ioan-Dragoş Tudorache (May 9, 2023), [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA\\_IMCOLIBE\\_AI\\_ACT\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf) accessed 25 May 2023.

conjunctive. The use of the words ‘such as’ followed by commas to separate the capabilities indicates that an AI system that simply performs image and speech recognition (and isn’t capable of pattern detection, question answering etc.) would still qualify as a GPAI under this definition. The second part of the definition, which states that a GPAI may be used in a plurality of contexts does not resolve this issue, as the use of the word ‘may’ means that being used in a plurality of contexts is not an essential condition to qualify as a GPAI, but merely a possibility.<sup>75</sup>

Most of the dangers associated with GPAIs as outlined in the previous pages would not apply to an AI system that simply does image and speech recognition, and has no other capabilities. The same would be true of an AI algorithm that only does translation and nothing else. And yet both would be considered as GPAIs under this definition. Thus the definition appears to be over-inclusive, and significantly so.<sup>76</sup> As such, it deviates from the approach AI experts have taken when defining GPAIs.<sup>77</sup>

The portions on General Purpose AI systems in the AI Act are chiefly captured under Articles 4a-4c, though there are other Articles in the Act that apply to GPAIs.

Article 4a(1) states that GPAI systems shall only comply with requirements given under Article 4b, and no other requirements mentioned in the AI Act (with the exception of provisions on prohibited AI practices under Article 5, transparency obligations under Article 52, measures relating to AI regulatory sandboxes under Article 53, and voluntary codes of conduct under Article 69).

Article 4b reveals the essence of the EU’s approach towards GPAIs. Unlike what might be initially assumed, this article doesn’t contain any specific requirements tailored to the functioning and risks of GPAIs. Instead, it posits requirements for when a GPAI system might be considered a high-risk AI system, and then states

---

<sup>75</sup> Hacker P, Engel A and Mauer M, ‘Regulating CHATGPT and Other Large Generative AI Models’ (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 5.

<sup>76</sup> Ibid.

<sup>77</sup> Gutierrez, Carlos Ignacio and Gutierrez, Carlos Ignacio and Aguirre, Anthony and Uuk, Risto and Boine, Claire and Franklin, Matija, A Proposal for a Definition of General Purpose Artificial Intelligence Systems (October 5, 2022) <https://ssrn.com/abstract=4238951>, accessed 25 May 2023.

that all the requirements in the act that apply to high-risk AIs shall apply to that GPAI system.

Essentially, if a GPAI system may be used as a high-risk AI system or as a component in a high-risk AI system, it shall have to comply with the obligations that apply to high-risk AI systems (Article 4b(1)).

Article 4c provides certain exceptions to Article 4b. 4c(1) states that Article 4b doesn't apply if the provider of the GPAI explicitly excludes all high risk uses in the instructions accompanying the GPAI. However, 4c(2) clarifies that this exclusion must be made in good faith, and if the provider has sufficient reason to consider that the GPAI may be misused, the provider will not be able to claim the exception of 4c(1), and will have to comply with the requirements for high-risk AI systems under Article 6.

Before proceeding with the analysis, it's relevant to understand what the AI Act considers as a high-risk AI system. Article 6 of the Act, which contains classification rules for high-risk AI systems, refers to Annex III, where a list of AI-systems considered high-risk is provided, based on the use of the system. Included are biometric systems, AI systems being used in critical infrastructure like roads, supply of water, etc, AI systems used in education for admissions or evaluations, AI systems used for recruitment of workers, AI systems used by law enforcement, and more.

A GPAI like ChatGPT could easily be used for many of those purposes. As such, read together, Articles 4a, 4b and 4c demonstrate a fundamental misunderstanding of GPAI systems. The nature of these systems means that it is unlikely that any provider will be able to claim in good faith that they do not reasonably expect the system to be misused. And if that is so, they will not be able to claim an exception under 4c. Which means that the GPAI will be treated as a high-risk AI, if it may be used for high-risk purposes. And based on the previous chapter of this thesis, it's a logical assumption that every GPAI may be used for high-risk purposes.

Thus the end-result of these provisions would essentially be to categorize almost all (if not all) GPAIs as high-risk AI systems.

The pertinent question here is that if this categorization is necessarily a problem. If the requirements imposed on high-risk AI systems also happen to work for GPAIs successfully, then almost all GPAIs being considered high-risk AI systems is not an issue. But if they don't work, that means the AI Act fails to regulate GPAI systems in an effective manner.

To answer this question, an analysis of Article 9, that establishes a risk management system for high-risk AI systems, is relevant.

Article 9(2)(a) requires the provider of an AI system to identify and analyse on a continuous iterative basis all “*known and foreseeable risks most likely to occur to health, safety and fundamental rights*” related to the high-risk purpose of the AI system, and Article 9(2)(d) read with Article 9(4) requires the provider to adopt measures for elimination or mitigation of those risks.

This would require the developers of GPAI systems to analyse every possible risk associated with every possible application considered high-risk under the Act, a process that is likely to be both inordinately expensive and patently unfeasible.<sup>78</sup>

The lack of predictability in the use-case scenarios of GPAIs means that such an analysis would be largely hypothetical, and might constitute a huge waste of resources by the provider to ascertain dangers that might not come to exist at all for a host of reasons.<sup>79</sup>

If providers of GPAIs are unable to comply with the requirements of the AI Act, it opens them up to not just the penalties envisaged under the Act, but also civil liability claims under the proposed directive on AI liability<sup>80</sup>.

Further, compelling the provider to deploy mitigation measures based on this hypothetical analysis is likely to cause more issues. Beyond the fact that users have demonstrated an acute ability to bypass the guardrails deployed by developers on GPAI systems (as has been mentioned in the previous chapters of this thesis), such

---

<sup>78</sup> Hacker P, Engel A and Mauer M, ‘Regulating CHATGPT and Other Large Generative AI Models’ (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, pages 5 and 6.

<sup>79</sup> Ibid, page 6.

<sup>80</sup> European Commission, ‘Liability Rules for Artificial Intelligence’ (European Commission) [https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\\_en](https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en), accessed 25 May 2023.

mitigation measures can have complex knock-on effects on the AI system. For example, the guardrails deployed by the developers of Stable Diffusion to prevent users from generating images of children being subjected to violence do not allow the user to input a prompt with the words ‘child’ and ‘knife’ in the same sentence. But this restriction also prevents users from generating images of children cutting a piece of bread with a knife. And it is no surprise that users have found a way to bypass those guardrails anyway.<sup>81</sup>

If the developers of ChatGPT, for instance, in complying with the AI Act, analyse a wide variety of potential risks and then deploy guardrails to mitigate all those hypothetical risks, those guardrails could possibly restrict the capabilities of the system in such a significant manner so as to make it veritably useless for most tasks users would like to achieve with it. There have already been reports of user complaints in that regard due to OpenAI’s existing safeguards.<sup>82 83</sup>

The consequences of these provisions on competition, especially on SMEs (Small and Medium Enterprises) could have been dire if not for Article 55a(3), which states that Article 4b requirements “*shall not apply to micro, small and medium-sized enterprises*”.

Even so, corporations that do not qualify as SMEs, but aren’t large enough to absorb the costs of compliance with the AI act would be forced to concede the competition to some of the largest players in the market (viz. Microsoft, Google, Meta, Apple), who can afford to expend resources on such compliance.<sup>84</sup>

However, this exception for SMEs in Article 55a(3) results in a complicated set of considerations and consequences when it comes to open-source GPAIs.

---

<sup>81</sup> Goud B, ‘Game Changing AI - It’s Completely Free!’ (YouTube, 6 March 2023) <https://www.youtube.com/watch?v=HUB2tFI5rqM&t=572s> accessed 25 May 2023.

<sup>82</sup> Grogan R, ‘Are Safeguards Destroying Chatgpt?’ (Medium, 25 April 2023) <https://medium.com/the-generator/are-safeguards-destroying-chatgpt-ed60009a44da> accessed 25 May 2023.

<sup>83</sup> mkglass, ‘I Think They Are Dumbing down ChatGPT. Each Update Seems to Limit It’s Abilities.’ (Reddit, 13 December 2022) [https://www.reddit.com/r/OpenAI/comments/zl078z/i\\_think\\_they\\_are\\_dumbing\\_down\\_chatgpt\\_each\\_update/](https://www.reddit.com/r/OpenAI/comments/zl078z/i_think_they_are_dumbing_down_chatgpt_each_update/) accessed 25 May 2023.

<sup>84</sup> Hacker P, Engel A and Mauer M, ‘Regulating CHATGPT and Other Large Generative AI Models’ (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 6

### 3.1.1 Open Source GPAIs – A complicated problem

A piece of software is considered open-source when its source code is available for anyone to inspect, modify and enhance.<sup>85</sup>

A GPAI being open source would allow anyone to download the GPAI on their local systems (be it PCs or smartphones) and run or modify it without any restrictions that may come with access to a GPAI like ChatGPT which is not open-source.<sup>86</sup>

Open-source GPAIs serve an important function in the market, as they prevent monopolization of the technology by mega-corporations.<sup>87</sup>

Just like the internet brought about the third industrial revolution<sup>88</sup>, AI is set to bring about the fourth industrial revolution<sup>89</sup>. Imagine an internet where a small group of mega-corporations control what the people see and can access. A loose parallel would be the restricted Internet access in China, where the State is in complete control of the information available to citizens, and of how the citizens interact on online platforms.<sup>90</sup> The consequences for democracy and citizens' rights in a world with monopolistic control over the internet would have been disastrous.

---

<sup>85</sup> Opensource.com, 'What Is Open Source?' (Opensource.com) <https://opensource.com/resources/what-open-source#:~:text=Open%20source%20software%20is%20software.or%20%22application%22%E2%80%94wo rks.> accessed 25 May 2023.

<sup>86</sup> Chowdhury A, 'Is CHATGPT Open Source?' (VideoGamer.com, 15 March 2023) <https://www.videogamer.com/tech/is-chatgpt-open-source/> accessed 25 May 2023.

<sup>87</sup> Dickson B, 'How Open-Source Llms Are Challenging Openai, Google, and Microsoft' (TechTalks, 7 May 2023) <https://bdtechtalks.com/2023/05/08/open-source-llms-moats/> accessed 25 May 2023.

<sup>88</sup> Smith, B. (2019). The Third Industrial Revolution: Policymaking for the Internet. Science and Technology Law Review, 3. <https://journals.library.columbia.edu/index.php/stlr/article/view/3621> accessed 25 May 2023.

<sup>89</sup> Martin J, 'The Fourth Industrial Revolution: AI and Automation' (The Observer, October 2017) <https://theobserver-qjaa.org/the-fourth-industrial-revolution-ai-and-automation> accessed 25 May 2023.

<sup>90</sup> Wang Y, 'In China, the "great Firewall" Is Changing a Generation' (Human Rights Watch, 28 October 2020) <https://www.hrw.org/news/2020/09/01/china-great-firewall-changing-generation> accessed 25 May 2023.



Similarly, there have been calls by public policy experts of the need to pre-empt a GPAI monopoly<sup>91 92 93</sup>, based on fears that AI being controlled by a group of mega-corporations could lead to a dystopian future.<sup>94</sup>

There are a number of open-source GPAI models in proliferation today<sup>95</sup>, and there is hope that they will broaden the market for GPAI systems and prevent big-tech monopolization.<sup>96</sup> It has been demonstrated that open-source GPAI systems running on local machines can match the capabilities of ChatGPT<sup>97</sup>.

The definition of GPAIs in the AI Act includes open-source models. Thus, all the obligations discussed so far in the thesis would also be applicable to open-source GPAIs. And as stated above, those obligations could be crippling on the providers of these systems.

This is where Article 55a(3) comes in. If the providers of the open-source GPAIs qualify as SMEs, they do not need to satisfy the requirements imposed on high-risk AI systems.

However, open-source models come with a unique set of regulatory risks. Since these models can run on local machines (some of the popular open-source GPAIs today run on laptops and phones<sup>98</sup>) any guardrails or restrictions placed on the GPAI by the original provider can be removed. Once this modified GPAI is free of those

---

<sup>91</sup> Coyle D, 'Preempting a Generative AI Monopoly' (Project Syndicate, 21 February 2023) <https://www.project-syndicate.org/commentary/preventing-tech-giants-from-monopolizing-artificial-intelligence-chatbots-by-diane-coyle-2023-02> accessed 25 May 2023.

<sup>92</sup> Chakravorti B, 'Big Tech's Stranglehold on Artificial Intelligence Must Be Regulated' (Foreign Policy, 11 August 2021) <https://foreignpolicy.com/2021/08/11/artificial-intelligence-big-tech-regulation-monopoly-antitrust-google-apple-amazon-facebook/> accessed 25 May 2023.

<sup>93</sup> Wright R, 'Google Must Be Stopped Before It Becomes an AI Monopoly' (Wired, 23 February 2018) <https://www.wired.com/story/google-artificial-intelligence-monopoly/> accessed 25 May 2023.

<sup>94</sup> Priestley T, 'This Week in AI Is Already Leading to Dystopian Futures' (Medium, 25 March 2023) <https://medium.com/@theo/this-week-in-ai-is-already-leading-to-dystopian-futures-86cde9b44fb> accessed 25 May 2023.

<sup>95</sup> KDnuggets, 'Top Open Source Large Language Models' (KDnuggets, 14 September 2022) <https://www.kdnuggets.com/2022/09/john-snow-top-open-source-large-language-models.html> accessed 25 May 2023.

<sup>96</sup> Dickson B, 'How Open-Source LLMs Are Challenging Openai, Google, and Microsoft' (TechTalks, 7 May 2023) <https://bdtechtalks.com/2023/05/08/open-source-llms-moats/> accessed 25 May 2023.

<sup>97</sup> Raieli S, 'The Infinite Babel Library of LLMs' (Medium, 8 May 2023) <https://towardsdatascience.com/the-infinite-babel-library-of-llms-90e203b2f6b0> accessed 25 May 2023.

<sup>98</sup> Edwards B, 'You Can Now Run a GPT-3-Level AI Model on Your Laptop, Phone, and Raspberry Pi' (Ars Technica, 13 March 2023) <https://arstechnica.com/information-technology/2023/03/you-can-now-run-a-gpt-3-level-ai-model-on-your-laptop-phone-and-raspberry-pi/> accessed 25 May 2023.

restrictions, it can easily be proliferated all over the world through the means of torrents or piracy websites, evading regulatory control.

In an interview with Venturebeat, Sam Willison, an open-source developer, echoed these concerns –

*“OpenAI, for example, has extra filters and rules in place to prevent writing things like a Hitler manifesto, he explained. “But once you can run it on your own laptop and do your own additional training, you could potentially train a fascist language model — in fact, there are already projects on platforms like 4chan that aim to train ‘anti-woke’ language models,” he said.”<sup>99</sup>*

When GPAIs that can generate images or videos are brought into the discussion, the possibilities get even more dangerous.

Any kind of regulatory rules that require GPAI providers to build safeguards and mitigation measures into their products become irrelevant with open-source GPAIs. As such, open-source GPAI systems occupy a complex space in the discussion on regulation of AI systems.

The AI Act handles these complexities rather poorly. It tries to impose the same obligations that apply for non-open-source GPAI systems to open-source GPAI systems, placing providers of open-source GPAIs in an impossible position (in that they cannot possibly build mitigation measures into their product that can’t be removed or modified), but at the same time lets open-source GPAI systems completely off the hook if the provider is an SME.

### **3.2 GPAI Provisions in the AI Act EP Version**

Once the proposed AI Act draft reached the European Parliament, further critical changes began to be discussed. One of them was to have GPAIs like ChatGPT be made part of Annex III (which lists high-risk AI systems) as a residual category,

---

<sup>99</sup> Goldman S, ‘With a Wave of New LLMS, Open-Source AI Is Having a Moment - and a Red-Hot Debate’ (VentureBeat, 10 April 2023) <https://venturebeat.com/ai/with-a-wave-of-new-llms-open-source-ai-is-having-a-moment-and-a-red-hot-debate/> accessed 25 May 2023.

with an exception being provided if the AI-generated content had undergone human review, making a person or organization liable for it.<sup>100</sup>

However, the AI Act EP Version of May 9, 2023 does not contain GPAI systems as part of Annex III.

Interestingly, the AI Act EP Version does attempt to resolve the issues with the definition of GPAIs. It does so by not only introducing a new definition for GPAIs, but also introducing a definition for ‘foundation models’. The two new definitions are as follows-

*“ ‘general purpose AI system’ means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed” (AI Act -EP Version, Article 3(1d))*

*“ ‘foundation model’ means an AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks” (AI Act -EP Version, Article 3(1c))*

The AI Act EP Version connects these two definitions in Recital 60e as follows-

*“AI systems with specific intended purpose or general purpose AI systems can be an implementation of a foundation model, which means that each foundation model can be reused in countless downstream AI or general purpose AI systems.”*

As such, the Act envisions a foundation model as an engine. A general purpose AI system could be built using a foundation model as the core that powers it. An example would be Microsoft’s Bing AI, that is an implementation of OpenAI’s GPT-4.<sup>101</sup> Thus GPT-4 here would be the foundation model, while Bing AI would be a general purpose AI system.

The distinction can be important when distributing obligations. However, for the critical analysis in this thesis, the same critiques that were previously lobbed at the provisions related to GPAIs in the AI Act can now be transferred to the provisions

---

<sup>100</sup> Bertuzzi L, ‘AI Act: EU Parliament’s Crunch Time on High-Risk Categorisation, Prohibited Practices’ (www.euractiv.com, 7 February 2023) <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-parliaments-crunch-time-on-high-risk-categorisation-prohibited-practices/> accessed 26 May 2023.

<sup>101</sup> Mehdi Y (Bing, 14 March 2023) [https://blogs.bing.com/search/march\\_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4](https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4) accessed 25 May 2023.

relating to foundation models in the AI Act EP Version, as GPAI systems are simply implementations of foundation models, and thus all of the stated dangers of GPAI systems emanate from the foundation models.

The AI Act EP Version places strict obligations on the provider of a foundation model under Article 28(b). These obligations, broadly in line with the obligations for high-risk AI systems in the AI Act, require providers of foundation models to demonstrate mitigation and reduction of foreseeable risks to health, safety, fundamental rights, the environment, democracy and the rule of law (Article 28b(2)(a)), and assess the performance, predictability, safety, corrigibility and cybersecurity of their models during development (Article 28b(2)(c)), among other requirements.

Foundation models that are used in ‘Generative AI’, which is an AI system that generates content (viz. text, images, audio, videos) have to follow a further set of requirements that involve compliance with transparency obligations (Article 28b(4)(a)), ensuring that there are “*adequate safeguards against the generation of content in breach of Union law*” (Article 28b(4)(b)), and making available to the public a sufficiently detailed summary of the copyright-protected data in the training data used to build the model (Article 28b(4)(c)).

This set of requirements for generative AI applies to all the examples of GPAI systems given in this thesis (viz. ChatGPT, Midjourney, Google Bard) as they all come under the umbrella of generative AI<sup>102</sup>.

GPAIs have created an ecosystem where the development and usage of the AI system does not follow the simple producer-to-consumer chain.<sup>103</sup> The AI value chain starts with the providers of foundation models (like OpenAI with GPT-4), and moves to deployers, which could be businesses that build upon the foundation model to release an application that performs either a specific task or a host of tasks (for example, Microsoft has deployed GPT-4 into its product Bing, which acts as a

---

<sup>102</sup> Ortiz S, ‘What Is Generative AI and Why Is It so Popular? Here’s Everything You Need to Know’ (ZDNET, 5 May 2023) <https://www.zdnet.com/article/what-is-generative-ai-and-why-is-it-so-popular-heres-everything-you-need-to-know/> accessed 25 May 2023.

<sup>103</sup> Härlin T and others, ‘Exploring Opportunities in the Generative AI Value Chain’ (McKinsey & Company, 26 April 2023) <https://www.mckinsey.com/capabilities/quantumblack/our-insights/exploring-opportunities-in-the-generative-ai-value-chain> accessed 25 May 2023.

search assistant to users). Beyond that the AI value chain could also involve importers, distributors, and other kinds of operators of end user-facing applications.<sup>104</sup>

The AI Act EP Version recognizes the need to distribute responsibilities along the AI value chain, and attempts to do so under Article 28. The Article extends the obligations applicable to high-risk systems under Article 16 of the AI act to deployers/importers/distributors of high-risk AI systems if they qualify certain requirements (like making a substantial modification to the AI system or putting their name/trademark on a high-risk AI system). Further, it compels information-sharing up and down the value chain, which would be required by the entities in the value chain to comply with the AI Act.

The AI Act EP Version follows this distribution of responsibilities under Article 28 with extensive obligations on deployers of high-risk AI systems under Article 29, as well as the requirement that a deployer of a high-risk AI system must conduct a fundamental rights impact assessment before the deployment of the system (Article 29a). The Article 29a obligations require the assessment to at minimum include information like –

*“categories of natural persons and groups likely to be affected by the use of the system” (Article 29a(1)(c)), “the reasonably foreseeable impact on fundamental rights of putting the high-risk AI system into use” (Article 29a(1)(e)), and “a detailed plan as to how the harms and the negative impact on fundamental rights identified will be mitigated.” (Article 29a(1)(h)).*

Again, the wide variety of ways in which GPAIs are or may be used, many of those uses possibly unintended by the providers or deployers of the GPAI, means that the fundamental rights assessment requirement under Article 29a is likely to be onerous and difficult to implement, particularly due to the fact that fundamental rights are fuzzy and don’t translate well into concrete technical safeguards.<sup>105</sup>

---

<sup>104</sup> Kuspert S and others, ‘The Value Chain of General-Purpose AI’ (Ada Lovelace Institute, 10 February 2023) <https://www.adalovlaceinstitute.org/blog/value-chain-general-purpose-ai/> accessed 25 May 2023.

<sup>105</sup> Hacker P, Engel A and Mauer M, ‘Regulating CHATGPT and Other Large Generative AI Models’ (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 7.

It is to be noted that the exception under Article 55a of the AI Act – which states that requirements for GPAI systems laid down in Article 4b (which would lead to a GPAI system to be considered high-risk) do not apply to SMEs – seems to have been removed in the AI Act EP Version. The new set of provisions that apply to GPAI systems (Articles 28, 28b, 29, 29a) do not provide for an exception for SMEs. Article 29a(4) of the AI Act EP Version, which requires deployers to notify the national supervisory authority and conduct a consultation with relevant stakeholders in course of the fundamental rights impact assessment, excludes SMEs from this requirement of notification and consultation, which would suggest that the larger requirement to conduct a fundamental rights assessment **does** apply to SMEs. If that is indeed the case, the extraordinarily heavy regulatory burden of these provisions will be disastrous for the existence of SMEs in the GPAI space.

That said, overall, the changes in the AI Act EP Version aim to create a better-fitting set of obligations for GPAIs by recognizing the role and responsibilities of various actors along the AI value chain, and peculiarities of GPAI systems (viz. foundation models, generative AI etc.). However, the changes still fail in important ways. The obligations imposed continue to not be suited to the complex and variable usage of GPAIs, and there is no consideration for the special challenges posed by open-source GPAIs.

## 4. Remedies and Suggestions to Address the Critical Failings of the AI Act

### 4.1 Would the AI Act benefit from a regulatory focus on users?

Philipp Hacker, Andreas Engel and Marco Mauer attempt to resolve some of these issues in their working paper titled ‘Regulating ChatGPT and other Large Generative AI Models’<sup>106</sup> with a novel set of suggestions.

The authors critique the EU strategy for the regulation of GPAs (they refer to GPAs as LGAIMs – Large Generative AI Models) for its focus on the providers of these AI systems, and not on the users using these systems for high-risk applications.<sup>107</sup>

They argue that regulatory focus for GPAs should shift towards deployers and users-

*“While some general rules, such as data governance, non-discrimination and cybersecurity provisions, should indeed apply to all foundation models, the bulk of the high-risk obligations of the AI Act should be triggered for specific use cases only and target primarily deployers and professional users.”<sup>108</sup>*

The argument is rooted in the nature of GPAI systems. These AI systems can be deployed for a large variety of uses, some of which may be considered high-risk under the AI Act, and even in cases where the GPAI isn’t specifically deployed for a high-risk use, users may be able to put it to high-risk uses regardless of whether the original developers could have predicted such use, or have designed safeguards

---

<sup>106</sup> Hacker P, Engel A and Mauer M, ‘Regulating CHATGPT and Other Large Generative AI Models’ (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023.

<sup>107</sup> Ibid, page 8.

<sup>108</sup> Ibid.

to prevent such use (like the example in this thesis of hackers bypassing ChatGPT's safeguards to make it generate malware).

As such, a regulatory focus on deployers and users might open up a better way to regulate GPAI systems.

Since the AI Act EP Version does give ample consideration to deployers (as discussed above), the thesis will move forward with an examination of the proposals of the working paper regarding regulatory focus on users.

The paper advances the idea of separating 'professional users' from 'non-professionals users' with regards to the purpose the AI output is being used for. For example, an advertising company using a GPAI to generate pictures for an ad campaign would be a professional user, whereas a person asking a GPAI for recipes to cook at home would be a non-professional user. Further, it introduces the term 'recipient' for any entity that consumes a product offered by a user. For example, a person looking at an AI generated ad campaign is a recipient.<sup>109</sup>

Pinning obligations on users brings up two immediate issues. First, users are unlikely to have the insight or resources (whether it's the training data used to build the GPAI system, or other information related to the system) to enable them to comply with serious obligations under the AI Act. To resolve this the authors of the paper suggest forced collaboration and resource sharing between the providers of GPAI systems and other actors along the AI value chain. To address concerns of trade-secrets and IP rights, they suggest non-disclosure agreements and other mechanisms inspired by the US pretrial discovery system.<sup>110</sup>

Second is the issue of liability. If obligations are distributed all along the AI value chain, when injury is caused, who should be held liable for it? To take an example, let's say an advertising company builds an ad campaign with an AI generated video that shows a thief entering a house. The video was generated by an application named ViDream, a GPAI system built upon a foundation model named Codex. It just so happens that the face of the thief in the video matches a real person, who

---

<sup>109</sup> Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 9.

<sup>110</sup> Ibid, page 10.



suffers harm due to the similarity. In this case, who can the injured person hold liable for this harm? Is it the ad company (the professional-user) that released the ad? Or the deployer of the app ViDream, which generated it? Or is it the provider of the foundation model Codex, which was used to build the app?

The authors suggest holding the actors in the AI value chain joint and severally liable. This way, the injured person doesn't have to suffer from these parties shifting blame from one to the other, and the parties held liable by the injured person can seek reimbursement from the other actors along the AI value chain.<sup>111</sup>

The suggestion of placing regulatory focus on users holds several benefits. Firstly, instead of placing tremendously onerous obligations on providers of GPAI systems, who cannot reliably predict how their product might be used by downstream users, the obligations regarding high-risk uses are placed on the very users employing the GPAI for high-risk purposes. If a corporation wishes to use ChatGPT for the selection of candidates for a potential position, it's sensible to place the high-risk obligations on that corporation rather than on OpenAI (the provider of ChatGPT), and instead compel OpenAI to share the resources the corporation would require to comply with those obligations. The approach is likely to be far more effective than the one suggested in the AI Act, which, as discussed earlier in this thesis, is built upon hypotheticals and suffers from issues of feasibility.

The AI Act EP Version does actually include obligations for professional users. The definition of deployer in Article 3(4) of the AI Act EP Version states –

*“ ‘deployer’ means any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity”*

As such, professional users are included in the definition of deployer. However, as discussed above, the obligations imposed on deployers in the AI Act EP Version are excessively cumbersome, and requiring professional users to comply with the same obligations without distinction is thus problematic.

---

<sup>111</sup> Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 11.

The authors of the paper go on to suggest that GPAI models should be subjected to a staged release process, where only selected stakeholders and security researchers would have access to the model for an initial period of time.<sup>112</sup> This risk management approach, that involves the community, could act as a test bed for GPAIs, and help detect possible harms, discriminatory outputs, and other flaws before the AI system is released to the wider public.

## 4.2 Addressing the Content Problem of GPAIs

GPAIs are capable of generating harmful and fake content which could be used by malicious actors to spread propaganda, cause chaos through deceptive and outrageous content and be used for mass-manipulation of citizens to undermine democracy.<sup>113</sup>

The authors point out that the Digital Services Act (DSA) that is meant to tackle such content, does not cover content created by GPAI systems<sup>114</sup>, and that this regulatory loophole must be plugged and the DSA must be expanded to cover GPAI systems better.<sup>115</sup> They also suggest building a community of trusted flaggers who would act as a decentralized content monitoring team. These flaggers would try to get the GPAI to generate problematic content, and on success, notify the developers of the AI system, who would prioritize those notices and make changes to the GPAI to tackle the detected issues.<sup>116</sup> Further, involving the community in the active and continuous monitoring of these systems goes some distance in solving the feasibility issues that plague the heavy obligations imposed by the AI Act on developers of a GPAI system.

---

<sup>112</sup> Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 19.

<sup>113</sup> Brundage M and others, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation' (arXiv.org, 20 February 2018) <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf> accessed 25 May 2023, page 6.

<sup>114</sup> Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, pages 15 to 17.

<sup>115</sup> Ibid, page 20.

<sup>116</sup> Ibid.

They also suggest tackling discrimination at the development stage of GPAIs by curation of the training data, so as to prevent propagation of that discrimination throughout the AI value chain. This could be done via auditing of training data and complementing real-world data with synthetic data.<sup>117</sup> An example would be that real-world data on doctors might disproportionately contain the names and images of men. Changing the names and synthetically adding images of female doctors could help balance out that bias and result in a GPAI model that is less prone to generate discriminatory content along those lines.

The authors conclude by stating that they believe that a technology-neutral law might be better suited to dealing with the new and emergent risks posed by GPAI systems.<sup>118</sup>

---

<sup>117</sup> Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf> accessed 25 May 2023, page 19.

<sup>118</sup> Ibid, page 21.

## 5. Exploring Alternative Regulatory Approaches

### 5.1 Is a technology-neutral law the way forward?

The allure of a technology-neutral law is that it might be more capable of keeping pace with new technological developments in the GPAI space, and unlike a technology-specific regulation that focuses on the current state of the technology, would not become obsolete by transformative changes in the space.

However, the argument that new technologies are best regulated by a technology-neutral law has been challenged by other scholars.

Lyria Bennett Moses, in her research paper – ‘Recurring Dilemmas: The Law's Race to Keep Up With Technological Change’<sup>119</sup>, outlines that technology-neutral drafting of laws in areas of evolving technology is a practical impossibility<sup>120</sup>, and the only way to draft a law that is truly technology-neutral and capable of encapsulating all kinds of new technological developments would be to draft at the highest level of abstraction, at which point, instead of drafting a rule such as ‘all GPAI systems must go through a fundamental rights assessment’ one would arrive at a rule like ‘all must act so as to preserve human rights’, which is both vague and unlikely to help regulate GPAIs effectively.<sup>121</sup>

This thesis would like to posit that the debate between technology-specific regulation versus technology-neutral regulation reflects a more fundamental regulatory debate regarding legal certainty versus regulatory flexibility.

A technology-neutral regulation that does not specifically focus on AI would necessarily have to follow a goals-based approach to regulation rather than a rules-based one. And to achieve those goals, the regulation would need to create or

---

<sup>119</sup> Bennett Moses, Lyria, *Recurring Dilemmas: The Law's Race to Keep Up With Technological Change* (April 11, 2007). UNSW Law Research Paper No. 2007-21, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=979861](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=979861) accessed 25 May 2023.

<sup>120</sup> *Ibid*, page 56.

<sup>121</sup> *Ibid*, page 62.

empower an institution (or an ecosystem of institutions) capable of framing and changing rules to achieve those goals as the space the institution is regulating evolves.

A technology-specific direct regulation like the AI Act (coupled with the AI liability directive) attempts to bring legal certainty to the AI space. Obligations and rights are clearly stated in the provisions of the Act. However, it does try to make the regulation somewhat future-proof by introducing a certain amount of regulatory flexibility. For example, Recital 12c of the AI Act recognizes the fast-paced technological developments and emergent scenarios in the market when it comes to GPAIs, and proposes conferring implementing powers on the Commission so that it may adapt the application of the requirements under the Regulation for GPAI systems to keep up with those changes. The proposal finds expression in Article 4b(1), which states the implementing acts to be adopted by the Commission shall specify and adapt the requirements established in the AI Act for high-risk AI systems (under Title III Chapter 2) *“to general purpose AI systems in the light of their characteristics, technical feasibility, specificities of the AI value chain and of market and technological developments.”*

Further, Article 7 of the AI Act gives the Commission the power to adopt delegated acts to amend Annex III (which contains the list of high-risk AI systems) to add high-risk AI systems to the annex when a set of conditions specified in the AI Act are fulfilled.

There’s also a catch-all provision under Article 67(1) where if a market surveillance authority (a national authority of a Member state given certain powers and responsibilities under the AI Act) finds that a high-risk AI system is compliant with the AI Act, and yet *“presents a risk to the health or safety of persons or to fundamental rights,”* then the market surveillance authority will require the operator of the AI system to take all appropriate measures to ensure that the system no longer presents that risk when put into service or placed on the market. If the AI system has already been placed on the market, the operator will be required to withdraw or recall it without undue delay within the prescribed period.

However, these attempts at regulatory flexibility are minor compared to the rigidity that permeates the AI Act.

The EU, confronted with the great present and potential risks of AI<sup>122</sup> on one hand, and the fears of being left behind in the AI race<sup>123</sup> on the other, has drafted a regulation that is exacting in its obligations so as to curb the risks of AI, but also presents a well-defined set of rules to provide legal certainty to AI developers in hopes of spurring the development of AI in the European Union.

It's a complex effort that, in the opinion of the author of this thesis, fails in the regulation of GPAI systems due to an inability to recognize not just the unique nature of these systems, but more importantly, the speed at which transformative changes are taking place in the GPAI space.

The core issue isn't merely that GPAIs pose a wide and complex variety of dangers. The issue is that both the development of GPAIs (and AI in general) and the purposes they are used for are evolving at a rapid pace, and any regulation that attempts to rein in the dangers while providing legal certainty risks either being obsolete quickly, or having obligations so onerous that they would cripple the development of AI technologies. The European Union, it appears, has drafted a regulation that is at risk of both.

## **5.2 The Choice between Legal Certainty and Regulatory Flexibility**

Whether a particular regulation should focus on legal certainty or regulatory flexibility will naturally depend on the field being regulated. Dr. Christopher Decker, in his research paper 'Goals-based and Rules-based Approaches to Regulation'<sup>124</sup>, provides a set of relevant contextual factors that can help decide whether a regulation that tilts towards legal certainty (what he refers to as a rules-

---

<sup>122</sup> Makortoff K, "'Risks Posed by AI Are Real': EU Moves to Beat the Algorithms That Ruin Lives' (The Guardian, 7 August 2022) <https://www.theguardian.com/technology/2022/aug/07/ai-eu-moves-to-beat-the-algorithms-that-ruin-lives> accessed 26 May 2023.

<sup>123</sup> Ibid.

<sup>124</sup> Decker, Christopher, Goals-based and Rules-based Approaches to Regulation (May 2018). BEIS Research Paper Number 8 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/714185/regulation-goals-rules-based-approaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714185/regulation-goals-rules-based-approaches.pdf) accessed 25 May 2023.

based approach) or towards regulatory flexibility (in his words a goals-based approach)<sup>125</sup> best serves a community.

The thesis will consider those contextual factors applicable for GPAI systems to conclude what approach may best serve a regulation seeking to govern the space-

1. The timing of intervention and costs of each approach – This is the question of whether higher costs are incurred (by regulatees, regulators, and legislators) when a regulation is given content before the regulatees act (thus giving legal certainty), or after the regulatees have acted (thus giving regulatory flexibility and responsiveness). If it is likely that the enforcement actions required will be based on largely similar fact situations, then the approach of legal certainty results in lower costs. However, if the situations are likely to be heterogenous in nature (and thus demand more bespoke enforcement actions), regulatory flexibility is the more cost-effective approach.<sup>126</sup> The entire thrust of this thesis has been to demonstrate that GPAI systems are quickly advancing in their capabilities, and have varied and unpredictable use-cases. As such, the cost-effective way to regulate them would be an approach that favours regulatory flexibility.
2. Simplicity or complexity of the context – Rule-based approaches that advance legal certainty are best used in simple settings where the actions that must be subjected to regulatory oversight are recurring and well-understood. Complex settings where neither the regulatees nor their actions are homogenous are best governed by a goals-based approach that advances regulatory flexibility.<sup>127</sup> The GPAI ecosystem, in this context, is complex both in the variety of actors participating in it, as well as the kinds of actions they are and will be engaging in.
3. Nature of the risks regulated and potential for regulatory error – The rules-based approach fails to capture a wide variety of potential risky

---

<sup>125</sup> Decker, Christopher, Goals-based and Rules-based Approaches to Regulation (May 2018). BEIS Research Paper Number 8 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/714185/regulation-goals-rules-based-approaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714185/regulation-goals-rules-based-approaches.pdf) accessed 25 May 2023, page 28.

<sup>126</sup> Ibid, pages 28 and 29.

<sup>127</sup> Ibid, pages 29 and 30.

behaviours, while a goals-based approach is capable of encapsulating a large spread of such behaviours.<sup>128</sup> Again, as demonstrated before, GPAIs enable a plethora of risky behaviours.

4. Information conditions – Dr. Decker recommends a shift away from a rules-based approach when the regulator does not have full access to the information required to frame effective rules.<sup>129</sup> The rapid-evolution taking place in the GPAI space, and the ever-new uses these systems are being put to naturally place the EU at an informational disadvantage when drafting rules for future governance.
5. Degree of Innovation – Dr. Decker proposes that a goal-based approach can allow for innovation both in terms of production processes and compliance. A regulatee may be able to find a new and innovative way to comply with the same regulation without the requirement of a rule-change.<sup>130</sup> This suits the GPAI space, as the EU does wish to promote innovation in AI.<sup>131</sup>
6. Communication, shared understandings and predictability – Dr. Decker postulates that the predictability and legal certainty said to be provided by a rules-based approach might be overstated, and that a goals-based approach can also be predictable if there is a shared understanding of goals and concepts between the regulator and the community being regulated.<sup>132</sup>

In light of the dichotomy between legal certainty and regulatory flexibility that started this discussion, this last point is interesting and merits further exploration. The proposition is that a regulation that follows a goals-based approach, and hence depends on the regulatory agency for interpretation and enforcement of those goals, might still achieve predictability if the agency is

---

<sup>128</sup> Decker, Christopher, Goals-based and Rules-based Approaches to Regulation (May 2018). BEIS Research Paper Number 8 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/714185/regulation-goals-rules-based-approaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714185/regulation-goals-rules-based-approaches.pdf) accessed 25 May 2023, pages 30 and 31.

<sup>129</sup> Ibid, page 31.

<sup>130</sup> Ibid, page 32.

<sup>131</sup> European Commission, ‘Artificial Intelligence (AI)’ (Research and innovation) [https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai\\_en](https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai_en) accessed 26 May 2023.

<sup>132</sup> Decker, Christopher, Goals-based and Rules-based Approaches to Regulation (May 2018). BEIS Research Paper Number 8 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/714185/regulation-goals-rules-based-approaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714185/regulation-goals-rules-based-approaches.pdf) accessed 25 May 2023, page 35.



able to cultivate a common understanding in the community being governed regarding those goals.

Thus legal certainty and regulatory flexibility may not be as diametrically opposed as they are often seen to be. In fact, there is an argument to be made that the legal certainty currently offered by the provisions of the AI act is illusory, because the changes in the AI space are likely to quickly render many of those provisions obsolete, and when that happens, the EU will be compelled to draft new rules to effectively govern the space and mitigate the risks that emerge. In that situation, the legal certainty promised by the AI Act would disappear in the face of those new rules, retroactively rendering it an exercise in futility.

### **5.3 Regulating Stochastic Parrots that have Learned to Swim**

Going back to the characterization of GPAIs as stochastic parrots by Timnit Gebru and her fellow authors – since the release of their ground-breaking paper, the complexity of GPAIs and the dangers they are fraught with have only grown.

These stochastic parrots have learned to swim. In the near future, they may develop more unique and unpredictable capabilities, and be used in even more surprising ways.

The EU's approach, one of building a golden cage around these parrots, does not work when these parrots are showing new abilities every day. They may just phase through the bars and fly away.

This thesis suggests that the focus of the regulation should not thus be on finalizing concrete rules for direct regulation of GPAIs (and AI in general), but on empowering a regulatory agency to regulate the space. A well-setup regulatory agency could anticipate and adapt to new changes in the AI space, and amend the regulatory rules that apply to the entities within so as to keep pace with the blistering speed of AI development.

To continue the metaphor, the regulatory agency could be a seagull that's capable of following the parrots into the skies and the seas, and of reining them in when need be.

That does not mean throwing away all the obligations and rights contained within the AI Act. The problem is that building a regulatory framework by means of direct regulation imparts a rigidity to the framework that is wholly unsuited to the pace at which the AI space is changing and developing.

The same regulatory framework, if, rather than being codified as hard law, was drafted as a set of rules that a dedicated agency may add to, remove or adapt as it sees fit in order to regulate the emerging scenarios in the AI space, that would suit the purposes EU aims to achieve with the regulation far better.

Mistakes are costlier in ink than pencil. And when the canvas is changing with every stroke, it's much more prudent to wield a pencil than a pen.

The idea that administrative agencies could play an important role in regulation of fast-changing technology isn't new. Lyria Bennett Moses proposed the same in 2007, highlighting that when unforeseen circumstances give rise to new problems, an agency staffed with technical experts could revise outdated rules much more quickly than a legislature ordinarily could.<sup>133</sup>

Neither is it alien to the EU deliberations on the regulation of AI. In 2017, the European Parliament adopted a text on 'Civil Law Rules on Robotics' that deliberated the establishment of a designated EU agency for Robotics and Artificial Intelligence *"to provide the technical, ethical and regulatory expertise needed to support the relevant public actors, at both Union and Member State level, in their efforts to ensure a timely, ethical and well-informed response to the new opportunities and challenges"*.<sup>134</sup>

---

<sup>133</sup> Bennett Moses, Lyria, Recurring Dilemmas: The Law's Race to Keep Up With Technological Change (April 11, 2007). UNSW Law Research Paper No. 2007-21, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=979861](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=979861) accessed 25 May 2023, pages 68 and 69.

<sup>134</sup> European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics [https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html) accessed 25 May, Statement 17.

This was followed by a call for a European regulatory agency for AI in a 2019 text adopted by the European Parliament titled ‘A comprehensive European industrial policy on artificial intelligence and robotics’.<sup>135</sup>

The AI Act itself envisions a constellation of authorities that would be involved in the governance of the framework created by the Act.

With the exception of the Commission, perhaps the most important of these bodies is a European Artificial Intelligence Board established under Article 56 of the Act. The membership of the Board constitutes of representatives from Member States (Article 56(2)), and the Board is tasked with establishing a subgroup for stakeholders who will advise the Board on implementation of the Regulation (Article 56(3)). While the focus on involving stakeholders in the regulatory process is commendable, the powers and tasks of the Board (under Article 58) are largely of an advisory nature, and it is clear from the provisions in Article 58 and 58a that the Board will exist as an accessory to the Commission.

Further, the Act requires the establishment of notifying authorities and market surveillance authorities as national competent authorities by each Member State (under Article 59 of the AI Act). These authorities primarily have a role in the enforcement of the Act (under Article 63).

It is evident from the structuring and powers of these authorities that their ability to actually change or shape the contours of the AI Act itself is non-existent.

As such, they are unlikely to be able to resolve the flaws in the Act in regulation of GPAI systems through regulatory responsiveness. Charlotte Stix, in her thesis titled ‘Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance’, explores the possibility of an empowered regulatory agency for AI in the EU in considerable detail.<sup>136</sup>

---

<sup>135</sup> European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics [https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html), Statement 187.

<sup>136</sup> Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology [https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420\\_Stix\\_hf.pdf](https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420_Stix_hf.pdf) accessed 26 May 2023.

Stix proposes a set of four interdependent institutions to govern AI in the EU –

1. The coordinator institution – Stix sees this as an umbrella organization for coordination of disparate policy efforts, norms and actions across various Member States and institutions in the EU.<sup>137</sup> It is worth noting that the European Artificial Intelligence Board proposed in the AI Act might perform a similar role.
2. The analyzer institution – The role of this institution would be to identify gaps in regulatory and implementation efforts across Member States, as well as track new developments in the AI space.<sup>138</sup> While no clear parallel for this institution exists in the AI Act, the European Parliament’s think tank does perform a similar role in a more general manner.<sup>139</sup>
3. The developer institution – Once the analyzer institution has identified regulatory gaps, it would be the responsibility of the developer institution to draft clear proposals to fill those gaps.<sup>140</sup> The closest analogue to this in the AI Act would be the Commission, but the powers of the Commission to amend the provisions of the Act are strictly curtailed, hence it might not be able to effectively fill regulatory gaps that appear as a result of the hard-coded obligations in the AI Act.
4. The investigator institution – Stix envisions the investigator institution as a watchdog that will monitor various actors in the AI space (be it corporations, organizations or governments) and ensure their adherence to the AI governance framework.<sup>141</sup> In the AI Act, the market surveillance authorities have some of the powers and responsibilities of a watchdog (Article 63 of the Act). However, these are national authorities, limited by Member State borders, unlike the supranational institution Stix proposes.

---

<sup>137</sup> Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology [https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420\\_Stix\\_hf.pdf](https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420_Stix_hf.pdf) accessed 26 May 2023, page 178.

<sup>138</sup> Ibid, page 183.

<sup>139</sup> ‘Home: Think Tank: European Parliament’ (Home | Think Tank | European Parliament) <https://www.europarl.europa.eu/thinktank/en/home> accessed 26 May 2023.

<sup>140</sup> Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology [https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420\\_Stix\\_hf.pdf](https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420_Stix_hf.pdf) accessed 26 May 2023, page 186.

<sup>141</sup> Ibid, page 189.

This framework of institutions, while comprehensive and elaborate, has serious issues.

Firstly, having a cluster of institutions governing the AI space could create issues of poor coordination (the existence of a coordinator institution notwithstanding), and entail significant resource costs in terms of time and effort spent during inter-institutional communication. Stix acknowledges these issues.<sup>142</sup>

Secondly, it could give rise to inter-institutional politics and rivalry, and result in a power struggle between the regulatory bodies. This is already a documented problem in the EU's supranational institutions.<sup>143</sup> There is a related issue of the new institutions being caught in a power struggle with the Commission, which could seriously jeopardize the functioning of the regulatory framework.

Thirdly, it is possible that Member States might not show deference to the authority of a set of new institutions in the way the Commission's authority has come to be respected by them. Lack of adequate compliance by Member States could create new headaches for these institutions, and result in time-and-resource-consuming litigation.

Fourthly, while multiple regulatory institutions do result in a dilution of power that is important for the functioning of a healthy democracy, they also complicate the policy process, and hinder the ability of the citizens to understand the process and hold the policy-making institutions accountable.<sup>144</sup>

Lastly, it would require a titanic effort for the EU to implement these proposals, first in law and then in infrastructure and logistics, with considerable expenditure of time and resources. The EU's current approach is built upon years of whitepapers, discussions and negotiations. The 180-degree pivot required for these

---

<sup>142</sup> Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology [https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420\\_Stix\\_hf.pdf](https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420_Stix_hf.pdf) accessed 26 May 2023, page 192.

<sup>143</sup> Müller Gómez, Johannes. (2018). Inter-institutional power struggle. The role of the European Parliament and the European Council in the EU's institutional architecture [https://www.researchgate.net/publication/325603780\\_Inter-institutional\\_power\\_struggle\\_The\\_role\\_of\\_the\\_European\\_Parliament\\_and\\_the\\_European\\_Council\\_in\\_the\\_EU's\\_institutional\\_architecture](https://www.researchgate.net/publication/325603780_Inter-institutional_power_struggle_The_role_of_the_European_Parliament_and_the_European_Council_in_the_EU's_institutional_architecture) accessed 26 May 2023.

<sup>144</sup> Kelbel C, Navarro J and Neihouser M, 'Interinstitutional Relations in the EU: Is Too Much Complexity Bad for Democracy?' (RECONNECT, 13 November 2020) <https://reconnect-europe.eu/blog/interinstitutional-relations-eu/> accessed 26 May 2023.

proposals would significantly delay the creation and subsequent functioning of these institutions, all the while actors in the AI space are running wild without regulation.

#### **5.4 From a Purpose-Driven Paradigm of Regulation to a Purpose-Driven Paradigm of Regulation**

The heading is not a typo. The EU's paradigm of regulation started as, and still remains focused on the purpose an AI system serves. High-risk purposes invite stricter obligations.

This thesis contends that the regulation should indeed be purpose driven, but the EU is looking for the purpose in the wrong place. It's not the purposes of the AI that should direct regulation, but the purposes the EU intends to achieve with the regulation.

Based on the contents of the AI Act, those purposes are likely to include the protection of fundamental rights, the protection of livelihoods, and the safety and privacy of the EU citizens.

The EU can empower an agency to regulate AI, and impose upon it the responsibility to uphold these purposes while doing so.

The law created as such will not be technologically neutral, because the agency it empowers will not be tech-neutral, but focused on AI. However, it will bypass the Achilles-heel of technology-specific legislation, which is its inability to adapt to rapid developments.

This thesis proposes that the institutional framework for the regulation of AI should be created within the Commission itself. An empowered Commission with the mandate to shape and change the rules that govern the AI space could be just the right approach to govern this rapidly-evolving, exciting yet dangerous space. In addition, it also resolves the issues arising out of Stix's proposal for multiple new institutions.

The work already done on drafting specific obligations, and suggestions by researchers like Philipp Hacker, Andreas Engel and Marco Mauer could be used by the Commission to draft an initial set of rules, which would then be modified as new developments arrive.

National and subnational authorities could be set up (similar to how the current AI Act does) to distribute responsibilities and ensure effective governance.

As the field evolves, over time, the Commission decisions and the case law could establish the legal certainty required of any mature regulatory framework. The entirety of the common law paradigm is based on building a body of laws over time through precedent. And it works as a system across the world. Thus common law may be used as an inspiration for this process.

### Risks and Challenges

This approach comes with its own set of challenges.

The most serious one is that it empowers the Commission at the expense of the European Parliament and the Council, and might disturb the balance of power between the institutions, an important prerequisite for democracy.

This issue can be resolved by making the Commission answerable to the Parliament and the Council for its regulatory decisions on AI, and giving power to these bodies, subject to the fulfilment of appropriate conditions, to ex-post-facto reject a regulatory change made by the Commission. The European Parliament already holds the right to approve or dismiss the European Commission, and the Commission submits regular reports on its activities to the Parliament<sup>145</sup>, so Parliamentary oversight in regulation of AI could fit seamlessly within the roles that these institutions play in the EU. Further, the Parliament can take the Commission to court if it believes the Commission has acted against the spirit of EU Law. More broadly, the CJEU already serves to balance the interests of

---

<sup>145</sup> European Parliament, ‘Supervisory Powers&nbsp;’ (Supervisory powers) <https://www.europarl.europa.eu/about-parliament/en/powers-and-procedures/supervisory-powers#:~:text=The%20European%20Parliament%20has%20the%20right%20to%20approve%20and%20dismiss.appear%20before%20an%20EP%20hearing.> accessed 26 May 2023.

institutions and individuals within the EU and can continue in that role with respect to issues that arise in the regulation of AI.

Another possible risk here would be that the loss of legal certainty would discourage AI developers from establishing their ventures in the EU. As has been discussed previously, legal certainty and regulatory flexibility may not be opposed concepts. However, in a situation where the Commission designs and changes the rules, predictability would require a shared understanding, trust, and communication between the regulator and the community. If the community feels blindsided by the decisions and policy movements of the Commission, that would disincentivize innovation, and possibly result in the community migrating to more attractive legal frameworks outside the EU. As such, this approach places immense responsibility on the Commission, and its powers must be exercised with deliberation and caution, and in close communion with the stakeholders likely to be impacted by policy changes.

Lastly, empowering the Commission in this manner is likely to resurface the criticisms commonly lobbed at the institution, namely that the institution lacks democratic legitimacy<sup>146</sup>, that its procedures and spending lack transparency<sup>147 148</sup>, and that it has dodgy relationships with lobbyists<sup>149</sup>.

There have been long-standing complaints by Member States that the Commission is prone to proposing too many detailed rules, that do not adequately consider the specific situations in various Member States.<sup>150</sup>

These are solvable issues. Unlike the GPAI space, the functioning of the Commission is predictable and well known, and any AI regulation that grafts new regulatory powers onto the Commission could also impose obligations for

---

<sup>146</sup> Vesnic-Alujevic L and Nacarino RC, 'The EU and Its Democratic Deficit: Problems and (Possible) Solutions' (2012) 11 European View 63 <https://journals.sagepub.com/doi/pdf/10.1007/S12290-012-0213-7> accessed 26 May 2023.

<sup>147</sup> EUR-Lex, 'Democratic Deficit' (EUR-Lex) <https://eur-lex.europa.eu/EN/legal-content/glossary/democratic-deficit.html> accessed 26 May 2023.

<sup>148</sup> Barr C, 'EU Commission Expenses Highlight Lack of Transparency in Brussels' (The Bureau of Investigative Journalism, 1 June 2011) <https://web.archive.org/web/20161123201657/https://www.thebureauinvestigates.com/2011/06/01/analysis-eu-commission-expense-scandal-highlights-lack-of-transparency-in-brussels/> accessed 26 May 2023.

<sup>149</sup> Riekeles G, '[Column] Eu Lobbying Clean-up -&nbsp;What Happened to That?' (EUobserver, 18 May 2023) <https://euobserver.com/opinion/157048> accessed 26 May 2023.

<sup>150</sup> Grant C, 'What Is Wrong with the European Commission?' (Centre for European Reform, 27 June 2013) <https://www.cer.eu/insights/what-wrong-european-commission> accessed 26 May 2023.



transparency on the institution, at least when it comes to the regulation of the AI space. The issues with transparency in lobbying go beyond a specific regulation and thus must be dealt with in a larger context. Finally, the need to have the voice of Member States in the regulatory process is important, and could be expressed through direct representation of the Member States within the regulatory body of the Commission responsible for the AI governance framework. The European Artificial Intelligence Board, composed of representative of Member States, could serve as a blueprint for a department or body within the Commission.

## 6. Conclusion

The thesis has illustrated the unique and complex dangers posed by GPAI systems, ranging from amplification of a biased and discriminatory worldview, the development of malware and generation of malicious content like fake news, propaganda etc., to posing a threat to peoples' livelihoods and generating false but plausible information.

It has examined the EU's strategy to regulation of GPAIs by an analysis of GPAI-specific provisions in the AI Act and the AI Act EP Version, and concludes that the EU has failed to adapt its original regulation paradigm successfully to GPAIs. It points out issues in the EU's provisions relating to obligations that are likely to be extraordinarily burdensome on the developers of GPAIs, the lamentable focus on Providers and Deployers when most of the high-risk use-cases take place in the hands of the users (and might not be intentioned by the developers or deployers), and poor consideration of the complex possibilities resulting from SMEs developing open-source GPAIs.

The thesis has then explored solutions to these issues pitched by other scholars, and the merits and demerits of those solutions, and finally arrives at the question if GPAIs would be better regulated by a goal-oriented law that empowers a regulatory agency instead of enacting hard-coded obligations within the law itself. The thesis considers various methods of implementing such an alternative regulatory framework, and concludes that the most optimal option would be for a goal-oriented law to empower the Commission to create, adapt and change the rules that would apply to GPAI systems in particular, and AI in general. The thesis examines the risks and challenges of this approach, and suggests mitigation strategies.

It is possible that with the deadline of the European Parliament vote on the AI Act approaching (which is currently expected to take place in mid-June 2023)<sup>151</sup>, the

---

<sup>151</sup> European Parliament, 'Ai Act: A Step Closer to the First Rules on Artificial Intelligence: News: European Parliament' (AI Act: a step closer to the first rules on Artificial Intelligence | News | European Parliament, 11 May 2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial->

suggestions in this thesis as well as those made by other scholars may not be taken up by the EU. If that happens, the entire European Union might be saddled with an Act that might either stifle GPAI or be left behind by it.

However, the true peril of the EU's AI Act isn't that the harshness of the provisions of the Act will kill GPAI, because at most it will only cause GPAI development to take place outside the EU<sup>152</sup>, but that the rigidity of the Act will render it obsolete and inapplicable at a moment when regulation might most be needed.

And that regulation will be needed is not a matter of debate. The CEO of OpenAI, Sam Altman, has himself urged U.S Congress to regulate AI<sup>153</sup>, as has Sundar Pichai, the CEO of Google<sup>154</sup>.

When the people not only best placed to understand the future development of a technology, but also in the position to best benefit from the proliferation of that technology - are asking for it to be regulated, that should terrify the rest of us regarding what that future holds.<sup>155</sup>

---

[intelligence#:~:text=On%20Thursday%2C%20the%20Internal%20Market,7%20against%20and%2012%20abstentions.](#) accessed 26 May 2023.

<sup>152</sup> OpenAI CEO Sam Altman recently stated that OpenAI, if unable to comply with the AI Act, will pull out of the European Union - Reuters, 'OpenAI May Leave the EU If Regulations Bite - CEO' (Reuters, 24 May 2023) <https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/> accessed 26 May 2023.

<sup>153</sup> Kang C, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing' (The New York Times, 16 May 2023) <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html> accessed 26 May 2023.

<sup>154</sup> Prakash P, 'OpenAI's Sam Altman and Google's Sundar Pichai Are Now Begging Governments to Regulate the A.I. Forces They've Unleashed' (Fortune, 23 May 2023) <https://fortune.com/2023/05/23/openai-sam-altman-google-sundar-pichai-begging-governments-regulate-a-i/> accessed 26 May 2023.

<sup>155</sup> Hale E, 'AI Could Cause Nuclear-Level Disaster, Third of Experts Tell Poll' (Technology | Al Jazeera, 17 April 2023) <https://www.aljazeera.com/economy/2023/4/14/ai-could-cause-nuclear-level-catastrophe-third-of-experts-say> accessed 26 May 2023.

# Reference list / Bibliography

## Official Publications

European Union

Commission, ‘Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS’, COM (2021) 206 final.

DRAFT Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council, Brando Benifei & Ioan-Dragoș Tudorache (May 9, 2023), [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA\\_IMCOLIBE\\_AI\\_ACT\\_EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf)

European Parliament resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics [https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html)

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics [https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html)

General approach adopted by the EU Council on Dec. 6, 2022 (Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS) <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>

## Literature

Adiguzel T, Kaya MH and Cansu FK, 'Revolutionizing Education with AI: Exploring the Transformative Potential of CHATGPT' (2023) 15 Contemporary Educational Technology <https://www.cedtech.net/download/revolutionizing-education-with-ai-exploring-the-transformative-potential-of-chatgpt-13152.pdf>

Alam A, 'Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education' [2021] 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3) <https://ieeexplore.ieee.org/abstract/document/9697300>

Baecker C and others, 'Threats Provided by Artificial Intelligence That Could Disrupt the Democratic System' (OPUS 4 | Threats provided by artificial intelligence that could disrupt the democratic system, 1 January 1970) <https://opus4.kobv.de/opus4-fhbrb/frontdoor/index/index/docId/2898>

Bender EM and others, 'On the Dangers of Stochastic Parrots' [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610 <https://dl.acm.org/doi/10.1145/3442188.3445922>

Bennett Moses, Lyria, Recurring Dilemmas: The Law's Race to Keep Up With Technological Change (April 11, 2007). UNSW Law Research Paper No. 2007-21, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=979861](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=979861)

Biswas SS, 'Potential Use of Chat GPT in Global Warming - Annals of Biomedical Engineering' (SpringerLink, 1 March 2023) <https://link.springer.com/article/10.1007/s10439-023-03171-8>

Borji A, 'A Categorical Archive of CHATGPT Failures' (arXiv.org, 3 April 2023) <https://arxiv.org/abs/2302.03494>

Brundage M and others, 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation' (arXiv.org, 20 February 2018) <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>

Decker, Christopher, Goals-based and Rules-based Approaches to Regulation (May 2018). BEIS Research Paper Number 8

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/714185/regulation-goals-rules-based-approaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/714185/regulation-goals-rules-based-approaches.pdf)

Future of Life Institute, 'General Purpose AI and the AI Act' (The Artificial Intelligence Act, May 2022) <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf>

Gutierrez, Carlos Ignacio and Gutierrez, Carlos Ignacio and Aguirre, Anthony and Uuk, Risto and Boine, Claire and Franklin, Matija, A Proposal for a Definition of General Purpose Artificial Intelligence Systems (October 5, 2022) <https://ssrn.com/abstract=4238951>

Hacker P, Engel A and Mauer M, 'Regulating CHATGPT and Other Large Generative AI Models' (arXiv.org, 12 May 2023) <https://arxiv.org/ftp/arxiv/papers/2302/2302.02337.pdf>

Li Z, 'The Dark Side of CHATGPT: Legal and Ethical Challenges from Stochastic Parrots and Hallucination' (arXiv.org, 21 April 2023) <https://arxiv.org/abs/2304.14347>

Miroshnichenko A, 'Ai to Bypass Creativity. Will Robots Replace Journalists? (The Answer Is "Yes")' (MDPI, 23 July 2018) <https://www.mdpi.com/2078-2489/9/7/183>

Müller Gómez, Johannes. (2018). Inter-institutional power struggle. The role of the European Parliament and the European Council in the EU's institutional architecture [https://www.researchgate.net/publication/325603780\\_Inter-institutional\\_power\\_struggle\\_The\\_role\\_of\\_the\\_European\\_Parliament\\_and\\_the\\_European\\_Council\\_in\\_the\\_EU's\\_institutional\\_architecture](https://www.researchgate.net/publication/325603780_Inter-institutional_power_struggle_The_role_of_the_European_Parliament_and_the_European_Council_in_the_EU's_institutional_architecture)

Roosevelt, Kermit III, "Certainty versus Flexibility in the Conflict of Laws" (2019). Faculty Scholarship at Penn Carey Law [https://scholarship.law.upenn.edu/faculty\\_scholarship/2029/](https://scholarship.law.upenn.edu/faculty_scholarship/2029/)

Sallam M, 'ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns' (2023) 11 Healthcare 887 <http://dx.doi.org/10.3390/healthcare11060887>

Smith, B. (2019). The Third Industrial Revolution: Policymaking for the Internet. *Science and Technology Law Review*, 3. <https://journals.library.columbia.edu/index.php/stlr/article/view/3621>

Stix, C. (2023). Towards Safe, Ethical and Beneficial Artificial Intelligence in the European Union and Beyond: A Multifaceted Framework for Governance. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Industrial Engineering and Innovation Sciences]. Eindhoven University of Technology [https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420\\_Stix\\_hf.pdf](https://pure.tue.nl/ws/portalfiles/portal/291353166/20230420_Stix_hf.pdf)

The Lancet Digital Health, ‘Chatgpt: Friend or Foe?’ (2023) 5 *The Lancet Digital Health* [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(23\)00023-7/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00023-7/fulltext)

Vemprala S and others, ‘ChatGPT for Robotics: Design Principles and Model Abilities’ (Microsoft, 20 February 2023) [https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT\\_Robotics.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2023/02/ChatGPT_Robotics.pdf)

Vesnic-Alujevic L and Nacarino RC, ‘The EU and Its Democratic Deficit: Problems and (Possible) Solutions’ (2012) 11 *European View* 63 <https://journals.sagepub.com/doi/pdf/10.1007/S12290-012-0213-7>

Wang D, Lu C-T and Fu Y, ‘Towards Automated Urban Planning: When Generative and CHATGPT-like AI Meets Urban Planning’ (arXiv.org, 8 April 2023) <https://arxiv.org/abs/2304.03892>

Zarifhonorvar A, ‘Economics of Chatgpt: A Labor Market View on the Occupational Impact of Artificial Intelligence’ [2023] *SSRN Electronic Journal* [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4350925](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4350925)

ZEMČÍK MgrT, ‘A Brief History of Chatbots’ [2019] *DEStech Transactions on Computer Science and Engineering* 14 [https://www.researchgate.net/publication/336734161\\_A\\_Brief\\_History\\_of\\_Chatbots](https://www.researchgate.net/publication/336734161_A_Brief_History_of_Chatbots)

Zhuo TY and others, ‘Exploring Ai Ethics of CHATGPT: A Diagnostic Analysis’ (arXiv.org, 22 February 2023) <https://arxiv.org/abs/2301.12867>

## Online sources

Abraham TM, 'I Got to Try GPT-4's Multimodal Capabilities and It's Quite Impressive! A Quick Thread of Examples...Let's Start out with Solving a CAPTCHA, No Big Deal' (Twitter, 16 March 2023)

<https://twitter.com/iScienceLuvr/status/1636479850214232064>.

Ankita, 'How to Bypass CHATGPT Filter Restrictions' (MLYearning, 9 May 2023) <https://www.mlyearning.org/how-bypass-chat-gpt-restrictions/>.

Barr C, 'EU Commission Expenses Highlight Lack of Transparency in Brussels' (The Bureau of Investigative Journalism, 1 June 2011)

<https://web.archive.org/web/20161123201657/https://www.thebureauinvestigates.com/2011/06/01/analysis-eu-commission-expense-scandal-highlights-lack-of-transparency-in-brussels/>.

Bayer M, 'EU Task Force to Review Chatgpt' (CIO, 24 April 2023)

<https://www.cio.com/article/474610/eu-task-force-to-review-chatgpt.html>.

be my eyes, 'Introducing Our Virtual Volunteer Tool for People Who Are Blind or Have Low Vision, Powered by OpenAI's GPT-4' (Be My Eyes - See the world together) <https://www.bemyeyes.com/blog/introducing-be-my-eyes-virtual-volunteer>.

be my eyes, 'The Story about Be My Eyes' (The story about Be My Eyes)

<https://www.bemyeyes.com/about>.



Bertuzzi L, 'AI Act: EU Parliament's Crunch Time on High-Risk Categorisation, Prohibited Practices' (www.euractiv.com, 7 February 2023)

<https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-parliaments-crunch-time-on-high-risk-categorisation-prohibited-practices/>.

Bertuzzi L, 'EU Council Presidency Pitches Significant Changes to AI Act Proposal' (www.euractiv.com, 1 December 2021)

<https://www.euractiv.com/section/digital/news/eu-council-presidency-pitches-significant-changes-to-ai-act-proposal/>.

Chakravorti B, 'Big Tech's Stranglehold on Artificial Intelligence Must Be Regulated' (Foreign Policy, 11 August 2021)

<https://foreignpolicy.com/2021/08/11/artificial-intelligence-big-tech-regulation-monopoly-antitrust-google-apple-amazon-facebook/>.

Chowdhury A, 'Is CHATGPT Open Source?' (VideoGamer.com, 15 March 2023)

<https://www.videogamer.com/tech/is-chatgpt-open-source/>.

Coyle D, 'Preempting a Generative AI Monopoly' (Project Syndicate, 21 February 2023)

<https://www.project-syndicate.org/commentary/preventing-tech-giants-from-monopolizing-artificial-intelligence-chatbots-by-diane-coyle-2023-02>.

Dickson B, 'How Open-Source LLMs Are Challenging OpenAI, Google, and Microsoft' (TechTalks, 7 May 2023)

<https://bdtechtalks.com/2023/05/08/open-source-llms-moats/>.

Edwards B, 'With Stable Diffusion, You May Never Believe What You See Online Again' (Ars Technica, 6 September 2022)

<https://arstechnica.com/information-technology/2022/09/with-stable-diffusion-you-may-never-believe-what-you-see-online-again/3/>.

Elci A, 'AI Deepfakes and Disinformation Loom over Turkey's Election'

(euronews, 12 May 2023) <https://www.euronews.com/next/2023/05/12/ai-content-deepfakes-meddling-in-turkey-elections-experts-warn-its-just-the-beginning>.

EUR-Lex, ‘Democratic Deficit’ (EUR-Lex) <https://eur-lex.europa.eu/EN/legal-content/glossary/democratic-deficit.html>.

European Commission, ‘Artificial Intelligence (AI)’ (Research and innovation) [https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai\\_en](https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/key-enabling-technologies/artificial-intelligence-ai_en).

European Commission, ‘Liability Rules for Artificial Intelligence’ (European Commission) [https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence\\_en](https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en).

European Commission, ‘Regulatory Framework Proposal on Artificial Intelligence’ (Shaping Europe’s digital future, 29 September 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

European commission, ‘A European approach to artificial intelligence’ (Shaping Europe’s digital future, 26 January 2023) <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.

European Parliament, ‘Ai Act: A Step Closer to the First Rules on Artificial Intelligence: News: European Parliament’ (AI Act: a step closer to the first rules on Artificial Intelligence | News | European Parliament, 11 May 2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence#:~:text=On%20Thursday%2C%20the%20Internal%20Market,7%20a%20gainst%20and%2012%20abstentions>.

European Parliament, ‘Supervisory Powers (Supervisory powers) <https://www.europarl.europa.eu/about-parliament/en/powers-and-procedures/supervisory-powers#:~:text=The%20European%20Parliament%20has%20the%20right%20to%20approve%20and%20dismiss,appear%20before%20an%20EP%20hearing>.

Fall JG, 'How It Started:•3,713 on Twitter•0 on Discord•\$0 Revenue How It's Going:•87,903 on Twitter•1,501 on Discord•\$130 Revenue' (Twitter, 19 March 2023) <https://twitter.com/jacksonfall/status/1637459175512092672>.

Future of life institute, 'About' (The Artificial Intelligence Act, N/A) <https://artificialintelligenceact.eu/about/>.

Future of life institute, 'Developments' (The Artificial Intelligence Act, N/A) <https://artificialintelligenceact.eu/developments/>.

Gangwar A, '12 Cool Things You Can Do with Chatgpt' (Beebom, 29 March 2023) <https://beebom.com/cool-things-chatgpt/>.

Goldman S, 'With a Wave of New LLMS, Open-Source AI Is Having a Moment - and a Red-Hot Debate' (VentureBeat, 10 April 2023) <https://venturebeat.com/ai/with-a-wave-of-new-llms-open-source-ai-is-having-a-moment-and-a-red-hot-debate/>.

Goodin D, 'Hackers Are Selling a Service That Bypasses CHATGPT Restrictions on Malware' (Ars Technica, 8 February 2023) <https://arstechnica.com/information-technology/2023/02/now-open-fee-based-telegram-service-that-uses-chatgpt-to-generate-malware/>.

Goud B, 'Game Changing AI - It's Completely Free!' (YouTube, 6 March 2023) <https://www.youtube.com/watch?v=HUB2tFI5rqM&t=572s>.

GPT Blogs, 'CHATGPT: How Much Data Is Used in the Training Process?' (GPT Blogs, 9 February 2023) <https://gptblogs.com/chatgpt-how-much-data-is-used-in-the-training-process>

Grant C, 'What Is Wrong with the European Commission?' (Centre for European Reform, 27 June 2013) <https://www.cer.eu/insights/what-wrong-european-commission>.

Grogan R, 'Are Safeguards Destroying Chatgpt?' (Medium, 25 April 2023)  
<https://medium.com/the-generator/are-safeguards-destroying-chatgpt-ed60009a44da>.

Hale E, 'AI Could Cause Nuclear-Level Disaster, Third of Experts Tell Poll' (Technology | Al Jazeera, 17 April 2023)  
<https://www.aljazeera.com/economy/2023/4/14/ai-could-cause-nuclear-level-catastrophe-third-of-experts-say>.

Härilin T and others, 'Exploring Opportunities in the Generative AI Value Chain' (McKinsey & Company, 26 April 2023)  
<https://www.mckinsey.com/capabilities/quantumblack/our-insights/exploring-opportunities-in-the-generative-ai-value-chain>.

Heikkilä, M. (2022) The viral ai avatar app lensa undressed me-without my consent, MIT Technology Review. Available at:  
<https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent>  
<https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent>

Home: Think Tank: European Parliament (Home | Think Tank | European Parliament) <https://www.europarl.europa.eu/thinktank/en/home>.

Hu K, 'CHATGPT Sets Record for Fastest-Growing User Base - Analyst Note' (Reuters, 2 February 2023) <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.

Kang C, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing' (The New York Times, 16 May 2023)  
<https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html>.

KDnuggets, 'Top Open Source Large Language Models' (KDnuggets, 14 September 2022) <https://www.kdnuggets.com/2022/09/john-snow-top-open-source-large-language-models.html>.

Kelbel C, Navarro J and Neihouser M, 'Interinstitutional Relations in the EU: Is Too Much Complexity Bad for Democracy?' (RECONNECT, 13 November 2020) <https://reconnect-europe.eu/blog/interinstitutional-relations-eu/>.

Khan Academy, 'GPT-4 Khan Academy in Depth Demo' (YouTube, 14 March 2023) <https://www.youtube.com/watch?v=rnIgnS8Sug>.

Khan S, 'What's Fake and Not Fake in Turkey's Election' (Inkstick, 16 May 2023) <https://inkstickmedia.com/whats-fake-and-not-fake-in-turkeys-election/>.

Korot E and others, 'Will AI Replace Ophthalmologists?' (Translational Vision Science & Technology, 28 January 2020) <https://tvst.arvojournals.org/article.aspx?articleid=2760332>.

Küspert S and others, 'The Value Chain of General-Purpose AI' (Ada Lovelace Institute, 10 February 2023) <https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/>.

Makortoff K, "'Risks Posed by AI Are Real": EU Moves to Beat the Algorithms That Ruin Lives' (The Guardian, 7 August 2022) <https://www.theguardian.com/technology/2022/aug/07/ai-eu-moves-to-beat-the-algorithms-that-ruin-lives>.

Margaret Mitchell: Google Fires Ai Ethics Founder (BBC News, 20 February 2021) <https://www.bbc.com/news/technology-56135817>.

Martin J, 'The Fourth Industrial Revolution: AI and Automation' (The Observer, October 2017) <https://theobserver-qiaa.org/the-fourth-industrial-revolution-ai-and-automation>.

Mehdi Y (Bing, 14 March 2023)

[https://blogs.bing.com/search/march\\_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4](https://blogs.bing.com/search/march_2023/Confirmed-the-new-Bing-runs-on-OpenAI%E2%80%99s-GPT-4).

mkglass, 'I Think They Are Dumbing down ChatGPT. Each Update Seems to Limit It's Abilities.' (Reddit, 13 December 2022)

[https://www.reddit.com/r/OpenAI/comments/zl078z/i\\_think\\_they\\_are\\_dumbing\\_down\\_chatgpt\\_each\\_update/](https://www.reddit.com/r/OpenAI/comments/zl078z/i_think_they_are_dumbing_down_chatgpt_each_update/).

Muehmel K, 'What Is a Large Language Model, the Tech behind Chatgpt?' (Blog, 15 March 2023) <https://blog.dataiku.com/large-language-model-chatgpt>.

Mulgrew A, 'I Built a Zero Day Virus with Undetectable Exfiltration Using Only CHATGPT Prompts' (Forcepoint, 9 May 2023)

<https://www.forcepoint.com/blog/x-labs/zero-day-exfiltration-using-chatgpt-prompts>.

OpenAI (Our approach to AI safety, 5 April 2023) <https://openai.com/blog/our-approach-to-ai-safety>.

OpenAI, 'About' (OpenAI) <https://openai.com/about>.

Opensource.com, 'What Is Open Source?' (Opensource.com)

<https://opensource.com/resources/what-open-source#:~:text=Open%20source%20software%20is%20software,or%20%22application%22%E2%80%94works..>

Ortiz S, 'What Is Generative AI and Why Is It so Popular? Here's Everything You Need to Know' (ZDNET, 5 May 2023) <https://www.zdnet.com/article/what-is-generative-ai-and-why-is-it-so-popular-heres-everything-you-need-to-know/>.

Pirnay E, 'We Spoke to People Who Started Using CHATGPT as Their Therapist' (VICE, 27 April 2023) <https://www.vice.com/en/article/z3mnve/we-spoke-to-people-who-started-using-chatgpt-as-their-therapist>.

Prakash P, 'OpenAI's Sam Altman and Google's Sundar Pichai Are Now Begging Governments to Regulate the A.I. Forces They've Unleashed' (Fortune, 23 May 2023) <https://fortune.com/2023/05/23/openai-sam-altman-google-sundar-pichai-begging-governments-regulate-a-i/>.

Priestley T, 'This Week in AI Is Already Leading to Dystopian Futures' (Medium, 25 March 2023) <https://medium.com/@theo/this-week-in-ai-is-already-leading-to-dystopian-futures-86cde9b44fb>.

Raieli S, 'The Infinite Babel Library of LLMs' (Medium, 8 May 2023) <https://towardsdatascience.com/the-infinite-babel-library-of-llms-90e203b2f6b0>.

Reuters, 'OpenAI May Leave the EU If Regulations Bite - CEO' (Reuters, 24 May 2023) <https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/>.

Riekeles G, '[Column] Eu Lobbying Clean-up - What Happened to That?' (EUobserver, 18 May 2023) <https://euobserver.com/opinion/157048>.

Scott G, 'Last Night I Made a Website That Uses GPT-4 to Code Any Arcade Game You Can Think of and Let You Play It Instantly. Here's a Demo of the Infinite Arcade. If People like It, I'll Publish the Site Later Today or Tomorrow!' (Twitter, 16 March 2023) <https://twitter.com/thegarrettscott/status/1636477569565335553>.

Sharma S, 'EU Closes in on Ai Act with Last-Minute CHATGPT-Related Adjustments' (Computerworld, 28 April 2023) <https://www.computerworld.com/article/3695009/eu-closes-in-on-ai-act-with-last-minute-chatgpt-related-adjustments.html>.

Sparrentak KV, 'Parliamentary Question: Lensa Image-Generation App: Concerns Regarding Bias, Harassment, Privacy, and EU and National Security: E-

004128/2022: European Parliament' (European Parliament, 21 December 2022) [https://www.europarl.europa.eu/doceo/document/E-9-2022-004128\\_EN.html](https://www.europarl.europa.eu/doceo/document/E-9-2022-004128_EN.html).

Volpicelli G, 'CHATGPT Broke the EU Plan to Regulate AI' (POLITICO, 6 March 2023) <https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/>.

Wang Y, 'In China, the "great Firewall" Is Changing a Generation' (Human Rights Watch, 28 October 2020) <https://www.hrw.org/news/2020/09/01/china-great-firewall-changing-generation>.

White MJ, 'The Dark Side of CHATGPT: Things It Can Do, but It Shouldn't' (Digital Trends, 16 April 2023) <https://www.digitaltrends.com/computing/bad-things-chatgpt-has-been-used-for/>.

Wright R, 'Google Must Be Stopped Before It Becomes an AI Monopoly' (Wired, 23 February 2018) <https://www.wired.com/story/google-artificial-intelligence-monopoly/> .