EU AI Act: A Comprehensive Analysis

Introduction:

The EU AI Act, a pioneer in the much-needed AI safety field, came into force at the start of August 2024 across all 27 EU member states. The enforcement of the provisions of the acts will be implemented in a phased approach extending till August 2027 (EU Artificial Intelligence Act, n.d. Implementation timeline.). Many organisations within and outside the EU that provide services in the EU are in an intense compliance checking phase, while some organisations are contemplating their stay and service in the EU. These practices from different companies make us question the reality or if there is one, of the global AI race (Sloane 2022) and Europe's stand in it. The essay aims to cover the different aspects of the AI governance mechanism in general and with regard to Large Language Models (LLMs) and use OpenAI's ChatGPT as a specific example of a General Purpose AI (GPAI) model that needs to comply with the EU AI Act.

EU AI Act provides rules and regulations to AI systems based on risk categorisation: Unacceptable risk, high risk, limited risk and minimal risk (EU Artificial Intelligence Act, 2024). ChatGPT's current free version using GPT 3.5 falls under the limited risk category. The essay will focus on the strengths and gaps of this specific use case and a general broader view. The EU AI Act's impacts on small and medium enterprises (SMEs), startups, science and research and user obligations are also examined. The essay will address the limitations and potential shortfalls faced by the governance mechanism such as regulating a fast-paced industry and vagueness/loopholes in addressing certain compliances including conformity assessments. With an additional focus on the root concepts of the act such as transparency, fairness and fundamental human rights, the essay will be able to state the conditions under which the EU AI Act may be successful in regulating LLMs such as ChatGPT.

Features of the EU AI Act:

The main framework of the AI governance mechanism is the risk-based categorisation of the AI systems and providing different regulations for the categories (EU Artificial Intelligence Act, 2024). The top-level risk is the unacceptable risk which is completely prohibited. The systems that come under these risks include systems that perform social scoring, biometric categorisation, attempt manipulative or deceptive techniques etc. The high-risk systems have the most strict compliances including performing an

internal or third-party conformity assessment before the products get released to the market. These systems may range from non-banned biometrics to systems used in education, employment, border management and law enforcement (EU Artificial Intelligence Act, n.d. Annex III). The limited-risk AI systems like Chatbots and deepfakes are subjected to lighter transparency such as informing the users that they are interacting with AI content and watermarking. Minimal risks such as AI-enabled video games and spam filters are unregulated.

The EU AI Act lies parallel with the existing GDPR, thereby improving the data protection laws in Europe. The act is comprehensive in nature and there are compliance checkers (EU Artificial Intelligence Act, n.d. Compliance Checker) available for the developers to detect which risk category they fall under. The act applies to both companies within the EU and for companies from other countries which provide services in the EU. Along with the main model provider, the downstream user who uses the model to create products such as fine tuning or using their API needs to comply with the act as well. However, the key point to note is that the act does not explicitly mention any regulation to the end-users like people who use ChatGPT daily (EU Artificial Intelligence Act, 2024).

LLMs and the AI Act:

Large Language Models are General-purpose AI (GPAI) models as they can perform a wide range of tasks from image/speech recognition to text generation and mathematical problem-solving. ChatGPT is such a GPAI, thereby needs to comply with the EU AI Act. GPAI models have different regulations based on whether they present a systemic risk i.e. if the model has high-impact capabilities evaluated based on appropriate technical tools/methods including indicators and benchmarks and/or if the computation used for its training measured in floating point operations (FLOPS) is greater than 10^25 (EU Artificial Intelligence Act, 2024). When they do have a systemic risk, the providers must conduct model evaluations, and adversarial testing, track and report serious incidents while also ensuring cybersecurity protections. If they do not have a systemic risk, the providers can only present documentation, and instructions for use, comply with the copyrights directory and publish a summary of training dataset content. Additionally, free and open-source models can just comply with copyrights and publish training data summaries without the rest.

The current free version of ChatGPT which uses the GPT 3.5 model comes under the limited risk category as it does not cross the 10^25 computation training limit. However, an advanced version of ChatGPT which uses GPT 4 is a model with systemic risks as it crosses the mentioned computation limits and possesses high capabilities. Looking at the free version of ChatGPT alone, OpenAI needs to only undergo limited regulatory actions compared to other strict rules.

Core concepts of the act:

Europe has always prioritized the data protection of its citizens. The EU AI Act is one such measure which ensures data protection in the world of data and datafication (Van Dijck 2014). This ensures the training datasets of the AI systems do not violate copyrights or steal personal data without the knowledge of the people. It also makes sure that datasets are representative, that is it is relevant to the model's task and contains diverse representations in (EU Artificial Intelligence Act, n.d. Article 10). Categorising systems used in spaces like border control and employment as high risk, the act makes sure to prevent or if not reduce the dataveillance that may take place in such places. For example, if an employer wants to use an AI system to monitor and evaluate their employees' performances, they need to establish a risk management system throughout the system's lifecycle, conduct data governance, submit technical documentation and instructions for use and mainly have human oversight along with other obligations. This ensures the concept of having humans before/on and after the loop of the AI system (Fabiano 2024) leading to respecting worker rights and creating fairness in the decision-making spaces.

The EU AI act ensures navigating two out of the three opacities mentioned by Jenna Burrell in (Burrell 2016). The opacity is intentional due to corporate or state secrecy and the opacity that arises from the characteristics of machine learning algorithms and the scale required to apply them usefully. For the corporate or state secrecy opacity, as suggested in the research, the EU AI act ensures to make the model's code available for scrutiny, either directly or through a 'trusted auditor'. This is done by providing the technical documentation and/or by conducting conformity assessments by a third party. For the opacity that arises from the black-box nature of the algorithms, the EU AI act completely prohibits them in certain areas such as social scoring and requires the others to be transparent and explainable.

Positives Pillars of the act:

AI safety is the latest focus of the technical field (Ahmed 2024). Frontier AI companies such as OpenAI, Google Deepmind and Anthropic address and acknowledge the future risks of advanced AI models and the crucial need for AI safety research (Anthropic, 2023). Particularly, Anthropic has a specific roadmap addressing the best, average and worst future-case scenarios of AI development and how if needed AI development should be stopped, and AI safety should become the only focus. These prominent visions by even the developers themselves emphasise the need for the governments to act on it. The EU successfully saw the vision and the need, thereby leading the global AI regulatory actions. It is an international strategy which involves the economic concerns around competitiveness and developing or retaining the global technological leadership as mentioned in (Sloane 2022).

The EU AI Act supports small and medium enterprises (SMEs) and startups by providing free and priority access to regulatory sandboxes where they can develop and test AI systems in a controlled environment under regulatory oversight. This will ensure the systems comply with the act, thereby reducing the need to face compliance costs. Additionally, the downstream users can lodge complaints on the models if they seem to not comply with the act i.e., startups which may use the frontier models for their products can ensure compliance by checking the provider's compliance details and complaining to the EU if suspicious.

In the context of LLMs, the high-level risk regulations only apply to certain models in the industry currently such as GPT 4, and Gemini Ultra which exceeds the 10^25 computation for training (Giskard, n.d.). This ensures there is enough space for innovation with smaller models. Free open-source models which are used for academic research or non-commercial purposes are exempted from the act if they do not possess unacceptable risk and use more than 10^25 for computation. This gives preference to innovations in AI research and development.

Gaps and limitations of the act:

While Europe's AI governance mechanism has a strong core base for developing a comprehensive global framework, it indeed faces many loopholes and shortfalls in the AI safety field. First and foremost, the whole act seems to mainly target the monopolies of the AI sector in the way it classifies the risks and their consequent regulations. There

are two reasons to assume this: the GPAI systemic risk classification is based on computation and the maximum obligations are to the commercial providers only.

In the former, the general-purpose AI models like GPT need to undergo model evaluations and adversarial testing only if the computation used for its training measured in floating point operations (FLOPS) is greater than 10^25. This mainly targets the AI monopolies which have access to higher computation powers. The idea that high computational models are the models with higher capabilities may be in line with the scaling law: performance increases with an increase in the number of parameters, training dataset size and training cost. However, since OpenAI's recent o1 model proved inference-time scaling: the model using more time to think is more effective in recent times, making the higher computation power for the training rule of the act irrelevant. Additionally, even computationally cheaper models can still provide systemic risk in terms of misinformation, bias, and hallucinations etc., Furthermore, it is unclear if the provider still needs to comply if they create a large model with systemic risks and distill it into smaller ones (Wachter 2024). Hence using FLOPS as a threshold for compliance does not effectively address the risks.

The second reason being the maximum obligations only to the commercial provider makes sense when seeing the exemptions of the act given to scientific research and personal AI development. This makes us wonder if an individual can develop a high-risk system in the name of research or non-professional activity, will they not face any immediate compliance fines? Does this paint a picture of bad capitalism and good consumerism? Such unanswered questions may be the major loopholes in the governance framework.

In the aspect of academic and scientific research, such an exemption may also indirectly contribute to the overreliance on AI systems in research and may further increase the AI's illusions of understanding among academics (Messeri 2024). There might be a primary focus on AI research, neglecting all other potentially more effective technologies which may lessen the number of innovations in the scientific field outside Artificial Intelligence. On the military and national security aspects, the AI systems are exempted from the EU AI Act. This may put Europe on the front line in terms of global politics but it also raises concerns about the internal usage of AI systems on whether there will be a similarity of usage like the algorithmic citizens in the United States (Cheney-Lippold 2016).

In a fast-paced industry, the regulatory actions will indeed be hard to keep up with. Regardless of the pace, there are still some gaps in the regulatory actions such as conformity assessments for high-risk systems and model evaluation tests for GPAIs. Only in the case of a biometrics-based system, if the provider does not apply harmonized standards/common specifications, there is an obligation for a third party to conduct the conformity assessment (Demetzou 2023). Otherwise, the providers themselves can conduct the assessments and inform the officials. This may lead to potential deceptions and failures similar to the situations of the self-regulating committees in the big tech. With systemic risks GPAIs, risk evaluations do not need to be public or submitted to competent authorities. The AI Office and national competent authorities only need to be informed about the corrective measures taken if serious incidents occur (Wachter 2024). This defeats the whole purpose of ensuring AI safety before any huge mishappening.

Similarly, in the limited risks category i.e., with chatbots or deepfakes, the providers should mostly only ensure that the user is interacting with an AI system. This does not address the potential societal and psychological risks that the bot can cause to the users. Instead of fixing the actual issue, the act only focuses on informing about it thereby prioritizing transparency over responsibility (Wachter 2024). It is the same case with environmental concerns where the act does not make sure to set standards on acceptable usage levels of energy consumption instead only obliges the provider to report the levels (Wachter 2024).

Furthermore, in the context of LLMs like ChatGPT, the EU AI act fails to recognise the stochastic parrot (Bender 2021) tendency of such models where humans may mistake LM output for meaningful text, in reality, it may not actually be meaningful due to its lack of understanding. This may also lead to potential risks such as over-reliance and trust in the AI sectors, particularly in sectors like health or education. At last, the act also does not address or explicitly state any obligations on the end-users for potential misuse of AI systems like the ChatGPT. For example, a recruiter can use ChatGPT in their daily use for the hiring process or someone with more malicious intent can use it for inventing dangerous weapons. In this case, if OpenAI has complied with all the regulations, the responsibility will be on the end-user for misusing it. The act does not explicitly specify any regulatory actions or checks for end-user misuse.

Global Consequences:

After the introduction of the EU AI Act, one of the leading AI developers Mistral AI struck a partnership deal with Microsoft and planned to remove the open access for its models (Digitizing Europe, 2024). While many data rights scholars supported the pioneer AI governance framework, there were many backlashes from the tech community. OpenAI recently released an AI future blueprint drawing similarities on how the automobile despite its origin in Europe bloomed in America (OpenAI, 2025). This raises concerns about whether there is an actual global AI race and if the US and China are the leaders with Europe significantly falling behind. While that is truly unclear at present but will face an answer soon if and when the AI economy starts rolling in. If there emerges an AI safety race, the EU might be the leader with its very early framework. The EU AI act also affects developing countries more as the companies from those regions may not be able to afford the high compliance costs in order to secure the European market.

Effective Conditions:

The LLMs like ChatGPT will be efficient and safe in the following conditions of the EU AI act: Implementing (Wachter 2024) suggestions of mandatorily having third-party conformity assessments/model evaluations instead of internal ones and changing transparency as the only accountability mechanism. Additionally, as she mentions "there is no such thing as unbiased data. AI models, systems, and their training and testing data should be assumed to be biased unless proven otherwise. This reversal can be accomplished by publishing the aforementioned testing results and actions undertaken to mitigate and prevent biases." This way the bias will be known and predictable. In terms of classifying systemic risks GPAI models, the FLOPS threshold should be lowered to include computationally smaller models that have similar systemic risks and can also use another criteria such as having a concrete number of end users, should be introduced.

The oversight mechanisms should be strengthened by involving more competent authorities for audits, thereby also creating new jobs. The obligations of the end-users should also be enforced by having traceability in the AI systems to hold the end-users accountable. The societal and psychological risks that systems like ChatGPT can cause such as generating harmful or misleading content and deception should be

addressed. Furthermore, critical attention to environmental risks needs to be given priority by assessing the environmental impacts as part of the audit (Wachter 2024). All these conditions will further enhance the act to make safer LLMs and other AI systems.

Conclusion:

Overall, the EU AI Act is a bold first step in the AI safety field by establishing a comprehensive governance mechanism for AI systems including Large Language Models such as ChatGPT. Through risk categorisation of the AI systems, the framework emphasises principles such as transparency, fairness, fundamental rights etc., The strengths of the act lie in its support for SMEs and startups and thorough address of the different AI systems. However, there are significant gaps in the act, particularly in addressing the computation of training-based risk classification, reliance on internal assessments, favouring transparency over responsibility and the absence of obligations for end-users. Such shortcomings can be addressed by a nuanced approach of refining the risk categorisation threshold and method, mandating third-party assessments, addressing societal and environmental risks and holding everybody who misuses the models accountable.

The AI governance mechanism faces many global consequences which may or may not cost its economy. Ultimately, the true success of the act in regulating AI systems for safer developments depends on its ability to adapt to the fast-evolving landscape of AI by creating safer innovations.

References

1. Ahmed, S., Jaźwińska, K., Ahlawat, A., Winecoff, A., & Wang, M. (2024). Field-building and the epistemic culture of AI safety. *First Monday*.
2. Anthropic. (2023, March 8). *Core Views on AI Safety: When, Why, What, and How.* https://www.anthropic.com/news/core-views-on-ai-safety
3. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
4. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*.
5. Cheney-Lippold, J. (2016). Jus Algoritmi: How the national security agency remade citizenship. *International Journal of Communication*, *10*, 22.
6. Digitizing Europe. (2024, February 26). *Microsoft-Mistral partnership and the EU AI Act.* https://www.kaizenner.eu/post/microsoft-mistral-partnership

7. EU Artificial Intelligence Act. (2024, May 30). *High-level summary of the AI Act.* https://artificialintelligenceact.eu/high-level-summary/
8. EU Artificial Intelligence Act. (n.d.). *Annex III: High-Risk AI Systems Referred to in Article 6(2).* https://artificialintelligenceact.eu/annex/3/
9. EU Artificial Intelligence Act. (n.d.). *Article 10: Data and Data Governance.* *https://artificialintelligenceact.eu/article/10/*
10. EU Artificial Intelligence Act. (n.d.). *EU AI Act Compliance Checker*. https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/
11. EU Artificial Intelligence Act. (n.d.). *Implementation Timeline.* *https://artificialintelligenceact.eu/implementation-timeline/*
12. Fabiano, N. (2024). AI Act and Large Language Models (LLMs): When critical issues and privacy impact require human and ethical oversight. *arXiv preprint arXiv:2404.00600*.
13. Giskard. (n.d.). *Regulating LLMs: What the EU AI Act means for Providers of Generative AI Systems.* *https://25507147.fs1.hubspotusercontent-eu1.net/hubfs/25507147/Giskard%20-%20Regulating%20LLMs_%20What%20the%20EU%20AI%20Act%20Means%20for%20Providers%20of%20Generative%20AI%20Systems%20-%20White%20paper.pdf*
14. Katerina Demetzou. (2023, November). *Conformity Assessments Under the proposed EU AI Act: A Step-By-Step Guide.* onetrust. https://fpf.org/wp-content/uploads/2023/11/OT-FPF-comformity-assessments-ebook_update2.pdf
15. Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49-58.
16. OpenAI. (2025, January 13). *OpenAI's Economic Blueprint*. https://openai.com/global-affairs/openais-economic-blueprint/
17. Sloane, M. (2022). Threading Innovation, Regulation, and the Mitigation of AI Harm: Examining Ethics in National AI Strategies. In *The global politics of artificial intelligence* (pp. 1-28). Chapman and Hall/CRC.
18. Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, *12*(2), 197-208.
19. Wachter, S. (2024). Limitations and loopholes in the EU AI Act and AI Liability Directives: what this means for the European Union, the United States, and beyond. *Yale Journal of Law and Technology*, *26*(3).