Unpacking competition concerns in Generative AI

Konstantinos Karalas[a*], Dimitrios Dechouniotis[a], and Fragkiskos Ninos[a]

[a]Hellenic Competition Commission, Athens, Greece

1. INTRODUCTION

Generative Artificial Intelligence (GenAI) is a type of Artificial Intelligence (AI) technique that can generate new realistic content by analyzing a variety of input data, including but not limited to text, images, audio, and videos [1]. GenAI has been widely studied since 2014 through Variational Autoencoders (VAEs) [2] or Generative Adversarial Networks (GANs) [3]. However, its rapid adoption has become possible due to recent advances in *(i)* computing power, *(ii)* Foundation Models (FMs) and Large Language Models (LLMs) which is a type of Deep Learning (DL) architecture for Natural Language Processing (NLP) tasks [4], and *(iii)* Multimodal AI which is an AI paradigm that can process multiple data types at the same time to achieve higher performance [5]. Famous chatbot assistants of big tech companies, such as Google's Gemini [6], Microsoft's Copilot [7], or OpenAI's ChatGPT [8] are based on advanced LLMs but provide user-friendly graphical interfaces enabling users without specialized technical knowledge to access them by asking for information in natural (and not programming) language, bringing democratization of data and AI in practice. One of them (ChatGPT) managed to become a valuable assistant for a wide range of tasks and thus one of the fastest-growing services ever, as shown in *Figure 1*. It is worth mentioning that the previous generation of the LLMs was shared openly with the research community, while the latest ones are not open-source (*e.g.*, GPT-4).

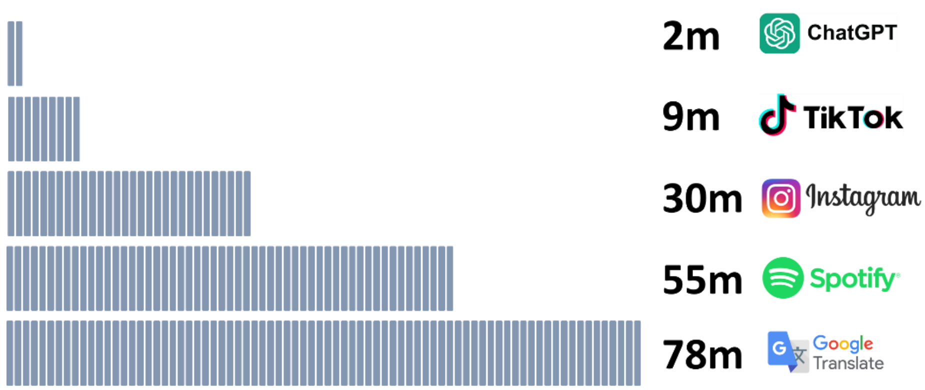

Figure 1: Time needed (in months) to reach 100M users for various applications (Source: UBS, Yahoo Finance).

GenAI systems fall under the broad category of Machine Learning (ML), as it is illustrated in *Figure 2*. Both learn from data and can significantly improve their performance by leveraging Big Data, but there is a substantial difference in their purpose and strategy. On the one hand, traditional ML models focus on processing data to unveil patterns and produce accurate predictions utilizing supervised, unsupervised, and reinforcement learning methods [9]. On the other hand, the goal of GenAI is to capture the underlying

* Further author information: (Send correspondence to K.K.)
E-mails: K.K.: kkaralas@epant.gr, D.D.: ddechouniotis@epant.gr, F.N.: fninos@epant.gr

data distribution and create new outputs (realistic facsimiles) that tend to mimic the real-world samples and human intelligence they have been trained on [10]. In this way, GenAI holds enormous potential to solve complex problems and create new value with economic impact for enterprises. More specifically, according to a recent report, the worldwide GenAI market worthed $40 billion in 2022 and is expected to reach approximately $1.3 trillion by 2032 [11]. Large companies in data-rich industries (*e.g.*, healthcare, finance, retail, manufacturing, energy, media, gaming) are the frontrunners in GenAI adoption and have already incorporated custom-built tools to revolutionize their processes in marketing, sales, product development, or service operations [12].

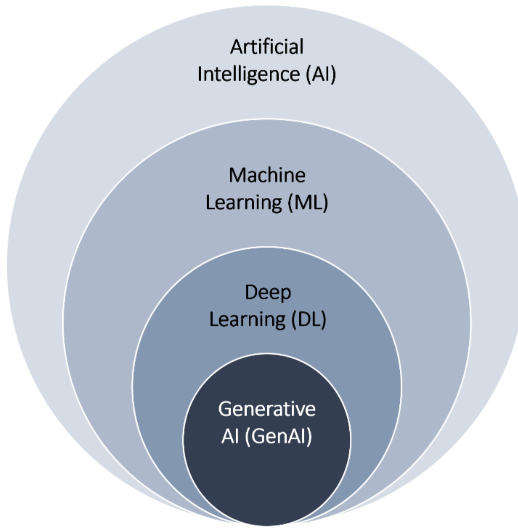

Figure 2: The relationship between AI, ML, DL, and GenAI.

Competition is also a privileged place for GenAI since there is an abundance of raw textual data, which can give unprecedented insights into complex questions leveraging FMs and LLMs. Nevertheless, GenAI can also introduce new risks regarding the enforcement of competition rules and the level of competition in the European Union's (EU) Market. For instance, many generative models may produce results that seem reliable but are in fact spurious or biased [13]. GenAI models can also ingest information that could be considered personal or copyrighted in their training data and thus promote it to their output, creating unique challenges for Privacy and Intellectual Property (IP) laws [14, 15]. Furthermore, since GenAI systems rely heavily on data for their performance, companies having access to large-scale datasets and powerful resources may lead to the creation of digital monopolies [16]. European Commission (EC) has already identified related risks in the data economy and tries to regulate AI systems by adopting rules such as the Data Act [17], Data Governance Act (DGA) [18], Digital Markets Act (DMA) [19], Digital Services Act (DSA) [20], AI Act [21], and the Ethics guidelines for trustworthy AI [22].

2. UNDERSTANDING THE GENERATIVE AI LIFECYCLE

Developing a comprehensive ML system with practical business value requires implementing an iterative set of steps, known as lifecycle [23]. Being a subset of ML, a GenAI project follows quite the same principles, albeit training a FM from scratch is a more challenging task that necessitates an expansion of the conventional ML lifecycle with large datasets, computational resources, and ML expertise. In general, the entire pipeline is a cyclic process and consists of 4 consecutive steps, *i.e.,* building, training, inference, and maintenance, as illustrated in *Figure 3*. Each component is unique, with its challenges, and affects the rest of the workflow. An organization needs to take action and define processes in each step to ensure a robust, scalable, and responsible GenAI system.

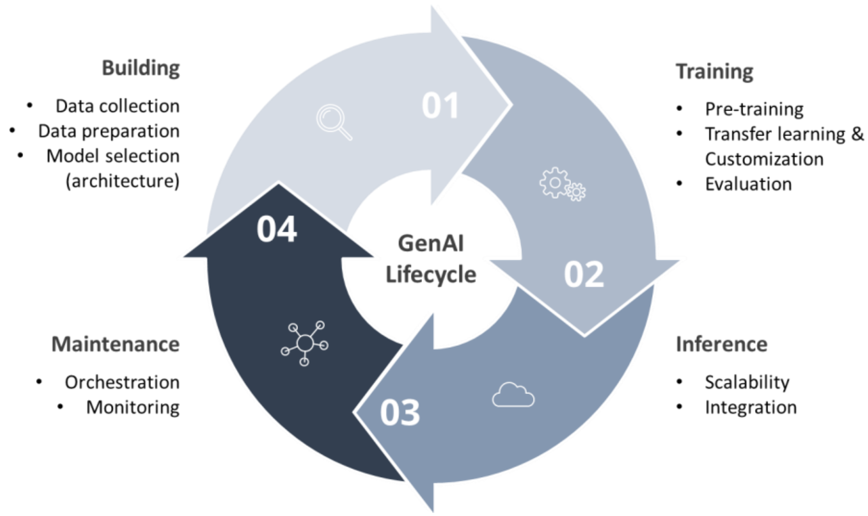

Figure 3: An overview of an end-to-end GenAI lifecycle.

The first crucial step for the effective development of a GenAI system is model building. According to data professionals (*e.g.*, data engineers, data scientists), this is a hard process since they must not only gather and analyze complex data from various sources but also gain business understanding to prepare them properly. There are 3 main parts in this stage:

- *Data collection*: GenAI models need to incorporate massive and varied data volumes (*e.g.*, from online news sources, wikis, articles, books, scraping from websites) representative of the domain they are intended to operate in. Collaboration with business experts is essential to understand the underlying business problem and collect appropriate data. It is noted that experts should engage, listen to, and work with stakeholders to estimate an acceptable baseline performance of the solution.
- *Data preparation*: It is a challenging and time-consuming process that plays a key role in data quality, a critical parameter for the development of a GenAI model. Tokenization (*i.e.*, breaking down text to individual tokens) is essential for FMs, since in this way they understand words and phrases in plain text. Applying practices such as data normalizing and cleaning (*e.g.*, removing irrelevant or noisy content), can lead to improved data quality and thus model performance.
- *Model selection*: The choice of model architecture depends on several factors (*e.g.*, the complexity of input data), and determines the model's underlying structure. Ultimately, it determines how the model learns from the raw data and generates synthetic content. There are multiple novel types of DL architectures for GenAI purposes, the most widely used are [24]:
    - *Autoregressive models* [25]: Used to generate sequences of data points by predicting one point at a time, conditioned on the previously generated point. Representatives include PixelRNN [26] for image, and WaveNet [27] for audio generation.
    - *Diffusion models* [28]: These models are commonly utilized in solutions offering high-quality images and videos. Their drawback is the slow and computationally intensive training process, which makes them inappropriate for real-time applications.
    - *Generative Adversarial Networks* (GANs) [3]: They can upscale images to a higher resolution quickly without losing detail, but the sample diversity is weak. This way, GANs are better suited for domain-specific data restoration (*e.g.*, medical images, satellite imagery).

- *Transformers* [29]: Transformers are designed to capture long-range sequential data non-sequentially and are considered state-of-the-art for NLP tasks. They are considered the fundamental building blocks of the well-known FMs and LLMs.
- *Variational Autoencoders* (VAEs) [2]: They are mostly used to generate images in a fast way. Nevertheless, their output is less detailed (i.e., blurry) compared to that of diffusion models.

After building the model, data professionals must train it to learn patterns from the data that have been prepared. In the case of GenAI, high computational costs are associated with the training of large-scale models which may include millions or even billions of parameters. This way, it is essential to incorporate specialized hardware resources, such as Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs) to accelerate the training process of large-scale FMs. There are 3 main parts in this stage:

- *Pre-training*: Developing a FM needs pre-training on large amounts of data. During this process the model learns general features and patterns (*e.g.*, captures grammar, syntax, context, and semantic relationships in the text).
- *Transfer learning & Customization*: In the context of DL and GenAI, several optimization techniques must be applied to ensure an acceptable model performance. For example, backpropagation is fundamental for LLMs to adjust model parameters (weights of neurons) and minimize the loss function, while optimization algorithms are employed to update the parameters in the network [30]. Additional training and fine-tuning methods to enhance the FM's performance on a specific task or domain (transfer learning) in a more complete fashion are also part of this process [31]. For instance, limiting the type of responses, training on a smaller dataset related to the target task, or condensing knowledge from a complex model into a simpler one (knowledge distillation).
- *Evaluation*: Model evaluation incorporates different metrics, such as perplexity for LLMs or task-specific metrics, to understand and measure the quality of the GenAI model. This way, data professionals can assess the strengths and weaknesses of the model objectively and transparently.

The next step in the process is the inference (deployment) of the model in a real-time modern production environment (*e.g.*, an application or a user interface). The goal here is to ensure that the trained model generates output based on the patterns it learned so that end users can enjoy accurate and reliable content, while ensuring data privacy and security. Typical tasks for a GenAI model include question answering, text generation, extraction, summarization, or language translation. There are 2 challenges involved in this stage:

- *Scalability*: Following the dominant cloud-native design principles, GenAI services should scale horizontally to handle varying loads and evolving requirements. Furthermore, they are often deployed with the microservices architectural concept to be flexible and modular.
- *Integration*: Towards offering a commercial GenAI service, it is necessary to develop various Application Programming Interfaces (APIs) that communicate with the target application and provide an integrated experience to the consumers.

The last step in the pipeline is to manage the model efficiently to continuously meet the users' requirements. For maintenance purposes, constant and iterative testing is crucial for refining and improving GenAI models to achieve customer satisfaction in various terms of Quality of Service (QoS). There are 2 main mechanisms in this stage:

- *Orchestration*: Orchestration strategies are used to schedule and manage robust workflows and scalable pipelines in an automated way. It is a common practice, first to containerize the GenAI model (*e.g.*, Docker) with all its dependencies, and then to orchestrate (*e.g.*, Kubernetes, OpenStack) and manage it in a distributed environment. Usually, the orchestration relies on well-known platforms of cloud providers, *i.e.*, Microsoft Azure, Amazon Web Services (AWS), or Google Cloud.
- *Monitoring*: Monitoring, logging, and tracking the model's performance is an ongoing process to detect issues promptly and gather valuable insights for improvement. If there are a lot of data changes, retraining can be applied to update the deployed model to the new circumstances. The ability to adapt quickly to evolving needs and market demands is essential (agile development). Companies that embrace agile methodologies and iterate rapidly can stay ahead of the competition.

Overall, one could say that the first two steps of the lifecycle concern "problem framing," while the last two steps concern "model operationalization", forming a complete framework to develop a functional GenAI application with genuine business value. It is worth mentioning that large generative models must be equipped with ethical considerations (*e.g.*, content filtering mechanisms, ethical guidelines) to ensure responsible, secure, and safe usage since the GenAI can produce biased, offensive, or inappropriate content. Towards addressing any ethical and law issue, the AI Act [21] and the Ethics guidelines for trustworthy AI documentation [22] provide guidelines on how the GenAI building blocks should meet three vital components: to be lawful, ethical, and robust.

3. KEY CONSIDERATIONS IN GENERATIVE AI

3.1 The role of data

GenAI is considered a data-driven innovation where the underlying data plays a pivotal role in the performance of the model. In other words, the more data is available, the better the model can learn and generate accurate and reliable content. Big Data, *i.e.*, exceptionally large datasets, is necessary to learn meaningful patterns for complex problems effectively and helps generalize better on unseen inputs and diverse target domains. Nevertheless, for a GenAI system, data quality is equally important, since a large volume of noisy data can lead to higher processing costs, biased models, unreliable results, and finally wrong decision-making. High quality can be assessed by various metrics and requires structured testing on the data. Typically, the following data quality dimensions should be examined for a robust model [32]:

- *Accuracy*: Relates to the correctness of the information, meaning that the data represent real-world, accurate, and reliable entities. Successful data governance can promote this data quality aspect.
- *Consistency*: Data must be uniform and coherent across various databases. It ensures that the data flows without loss between the architecture components of the organization.
- *Integrity*: Assures that digital information is stored and maintained in a database uncorrupted, while it can be accessed or modified only by authorized users.
- *Validity*: It is a measure of the reliability of data where value attributes should be aligned with the specific requirements and constraints of the domain according to predefined standards or rules.

Apart from the above conventional data quality metrics, for an operational GenAI system, there are many more dimensions of the data that should be assessed. First, since we live in a highly dynamic environment, data may quickly become outdated. Thus, constant data updates are necessary for GenAI applications to capture all the temporal changes and remain up to date. Furthermore, domain-specific curated context in training is considered essential to improve the performance of a real-world GenAI system according to the specific requirements of the target industry. Of course, diversity is also needed for a robust and reliable GenAI application to generalize. Finally, data privacy and security aspects are of paramount importance for a modern business application. Companies should enforce processes to protect sensitive or personally identifiable information from unauthorized access or other malicious activities. In parallel, they should apply standards to handle, process, and store data responsibly and comply with the latest legal and ethical guidelines.

The success of generative AI systems relies heavily on the careful curation, preprocessing, and utilization of relevant data. Companies and authorities cannot fully trust Big Data and must always evaluate GenAI model outputs critically due to potential biases, inaccurate information, or sensitive elements violating IP Rights (IPR). Ensuring that the data used for training is of high quality, representative, and aligned with the intended application domain contributes to the effectiveness and reliability of the resulting GenAI model.

3.2 The role of interoperability

Interoperability refers to the ability of two or more software systems to connect, communicate, and exchange information in a coordinated manner, work together, and use each other's functionalities seamlessly [33]. At the core of interoperability is the integration with existing workflows, applications, or infrastructure. Thus, it enables a GenAI component to be a part of a larger ecosystem and enhances its capabilities, rather than being an entity that is isolated from its surroundings. However, interoperability goes beyond the technical level for businesses and users. At the business level, interoperability facilitates inter- and cross-organizational collaboration between different stakeholders (*e.g.*, companies, non-profit organizations, research institutes) across the digital markets value chain promoting business exploitation opportunities since it enhances the system's capabilities by leveraging features from other connected services [34]. At the user level, interoperability ensures the seamless incorporation of GenAI models into applications, leading to a more integrated and user-friendly customer experience [35].

In this way, interoperability is related to modern data access, transmission, and portability [36]. Since GenAI models often rely on data from various sources, interoperability ensures smooth data exchange between different components, enabling GenAI applications to leverage diverse datasets and knowledge. For an effective competition, interoperability should involve the development and adoption of well-established industry standards. Standardization encourages inclusivity and consistency, making it easier for GenAI components to communicate and for companies to cooperate systematically. It is noted that EU Data Spaces pay special attention to interoperability by establishing a set of guiding principles [37], *e.g.,* the ISO 19941 Cloud Computing Interoperability and Portability [38] or the European Interoperability Framework (EIF) [39]. Both protocols aim to improve interoperability for data sharing, boost trust, and encourage investments. Moreover, interoperable systems are also flexible and adaptable to changes in technology and business needs. As a result, users can easily switch between components without facing any compatibility issues.

The lack of interoperability between components in the GenAI ecosystem can pose significant risks to competition [40]. In general, companies can benefit from a lack of interoperability by creating isolated environments for their products, which do not allow communication with the products or services of other organizations. Such actions may lead to fragmented ecosystems of lower performance and higher costs [41]. Additionally, the inability to exchange data seamlessly between different GenAI components might result in data silos, hindering the holistic utilization of information and reducing the effectiveness of the models. Overall, the lack of interoperability can be seen not only as a lack of choice for the end users but also as an obstacle to innovation since closed systems hinder the exchange of information and ideas between different players in the GenAI ecosystem [42, 43].

Addressing interoperability challenges is essential to foster a more competitive and collaborative GenAI community. To this end, a more interoperable ecosystem promotes open standards, encourages the use of common interfaces, prioritizes compatibility among diverse components, and supports effective stakeholder communication. Such initiatives can mitigate the risks associated with the lack of interoperability and support a more dynamic, transparent, and innovative GenAI landscape.

4. COMPETITION CONCERNS FROM GENERATIVE AI

In the future, various competition challenges are likely to emerge over one or more of the building blocks that GenAI relies on. Several types of potential competition concerns could arise in this context. However, the extent to which these concerns warrant intervention requires careful consideration of the key competition parameters in GenAI markets, the economic and regulatory realities, and an assessment on a case-by-case basis. Competition concerns can be broadly classified into the following categories:

A) Anticompetitive practices by dominant companies. Companies with a dominant position in markets that constitute essential inputs for GenAI could refuse or restrict the access of emerging downstream competitors to their data, essential hardware (*e.g.*, GPUs), foundation models, or services (*e.g.*, cloud computing). They could also leverage their dominant position by binding customers and end users into their proprietary solutions, for example through tying or bundling practices, thus strengthening their position in GenAI markets. These exclusionary practices could prevent emerging rivals from entering GenAI market or raise their costs. Last, dominant companies could impose unfair trading terms when providing access to their GenAI solutions to customers and end users.

Regarding the first two steps in the technical implementation of a GenAI system (*i.e.*, building and training), businesses that maintain large datasets collected from their users through the years (especially if they own online platforms) are possible to pursue monopolistic control over data to feed and optimize their model building and training processes exclusively (data monopoly) [47]. This way, startups and SMEs may have obstacles collecting or accessing large and diverse amounts of data to develop state-of-the-art GenAI models, particularly in sectors where data is more highly regulated, such as healthcare or finance[1].

Furthermore, pre-training necessitates access to powerful computational resources available through expensive on-demand cloud services offered by a limited number of big corporations. It is noted that the availability of open-source pre-trained models may be a key factor in deciding whether computational resources are a major barrier to entry in GenAI development [48].

Apart from that, dominant companies may attempt to lock-in their research workforce or protect their algorithms and pre-trained models with patents, which may in turn increase the risk of hindering

[1] Nevertheless, data bias remains a persistent threat, either for large or small companies. More specifically, when the training data are more representative of certain sample groups, the model will systematically favor those groups in its generated outputs leading to discriminatory results, which raises concerns related to fairness, inclusion, and equity. Consequently, addressing bias becomes vital to GenAI model development, as it can directly influence decision-making and thus business processes.

innovation and open-source initiatives by disrupting the balance between protecting IPR and fostering competition. The EU has tried for years to set some well-defined rules for the IPR of AI models and finally in 2023 the first version of the EU AI Act was published [21]. GenAI though, as a novel technology with distinctive characteristics and needs, is currently under continuous advancements to depict its operational and policy framework [49, 50].

The subsequent steps in implementing a GenAI system (*i.e.*, deployment, distribution) concern the exploitation of the models and are equally critical for effective competition. For instance, the lack of distribution channels and interoperability can create vendor lock-in issues, limiting consumers' ability to change providers, which gives a competitive edge to companies with operational GenAI models.

B) Algorithmic collusion. GenAI agents along with computer science could facilitate implicit unlawful agreements, even in the absence of direct human-to-human communication channels. In other words, such state-of-the-art technology offers opportunities to engage in algorithmic collusion making it hard for the competition authorities to uncover evidence [44, 45, 46].

C) Collaborations and acquisitions. Partnerships along with mergers and acquisitions (M&A) activity in the GenAI space should be closely monitored. Outright acquisitions or partnerships falling short of control could prevent or hobble the entry or expansion of emerging competitors, ultimately leading to decreased innovation and choice.

D) Other considerations. The absence of standardization in interfaces may lead to fragmented ecosystems, limited API integrations, and detached tools. This landscape could potentially impede collaboration among enterprises and creates barriers for new players. However, smaller companies aiming to build interoperable solutions of common standards and practices may find opportunities to gain a significant competitive advantage in the future, as stated in the Common EU Data Spaces [37]. Lastly, since access to GenAI is not strictly regulated, remarkable costs may arise for sophisticated technological tools. New entrants or SMEs may face challenges to meet all the necessary regulatory requirements and finally be excluded from the larger companies, not able to catch up.

Competition authorities should thoroughly study and understand the critical elements of the GenAI market value chain, *i.e.* computing resources, models, and data [51]. With this capacity, the competition bodies should investigate the long-term impact of established cloud providers, chip and graphic card manufacturers, and other involved industry sectors on the development of GenAI systems [52]. In parallel, the authorities should monitor the GenAI impact in several digital markets, which are early adopters of GenAI technology, such as search engines and online advertising. Addressing these competition issues requires a balance between fostering innovation and ensuring open competition and fair growth of startups and SMEs [53]. It is also important to strike the right balance between ex ante (through AI regulation and other tools, such as the DMA) and ex post intervention (through competition law tools). Regulatory frameworks, industry standards, open-source, and collaborative initiatives can play a crucial role in mitigating these challenges and promoting a healthy GenAI ecosystem.

5. CONCLUSION

There is no doubt that GenAI is a disruptive data-driven technology that promises the transformation of businesses across the entire value chain. It is then no surprise that enterprises, irrespective of size, are increasingly looking to GenAI to streamline their operations and decision-making. However, the businesses and the end users need to assess continually the potential negative impacts associated with irresponsible or unethical use of GenAI. In parallel, GenAI advancement offers new challenges and opportunities for the competition and antitrust authorities. This way, they need to adjust their processes on online market monitoring, dawn raids, and competition investigations to address the specific challenges posed by companies leveraging this technology, such as search discrimination, price steering, or tacit and explicit collusion. Furthermore, they should start actively examining the future effects of GenAI applications and develop new legal processes to mitigate anti-competitive strategies and promote the concepts of accountability, transparency, and confidentiality. Apart from updates in the legal treatment and regulation of GenAI, authorities should explore use cases to leverage the GenAI capabilities to boost efficiency in their internal functions, *e.g.* by developing document (case) search engines based on LLMs to support the needs of case handlers on fast information retrieval, extraction, and complex questions answering.

Integrating GenAI solutions with other digital tools or open-source intelligence can seamlessly enhance existing workflows and deliver significant added value. It is worth mentioning, that while GenAI can be a valuable tool, critical thinking on its outputs is essential. In the end, GenAI can assist existing processes by providing insights but not replacing human judgment. Towards this direction, competition and antitrust authorities should act on the three following pillars to mitigate risks of GenAI and at the same time ensure positive gains: *(i)* modern legislative/regulatory frameworks, including antitrust laws and competition policies, *(ii)* data-driven tools for market evaluation and pro-active cartel detection (*e.g.*, price discrimination analysis, collusion risk assessment, data screening for bid-rigging in public procurement [54, 55]), and *(iii)* AI tools for electronic discovery (e-discovery) of evidence and competition investigations (*e.g.*, semantic search, Technology-Assisted Review, chat-based data analytics, conceptual clustering). All these actions can be feasible under a single fundamental condition: close collaboration among all stakeholders (*i.e.*, EU, national competition authorities, regulatory control bodies, organizations, and academia). The collaboration should allow a common understanding and agreement on the competition concerns of the GenAI landscape. Subsequently, the development of comprehensive data governance and standardization of models along with effective regulations will foster innovation, protect users, and ensure a fair competition committed to the ethical and trustworthy application of AI on GenAI [51, 56].

REFERENCES

[1] Jovanovic, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, *55*(10), 107-112.

[2] Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

[3] Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, *35*(4), 3313-3332.

[4] Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y. H., ... & Jararweh, Y. (2024). Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*, *27*(1), 1-26.

[5] Suzuki, M., Nakayama, K., & Matsuo, Y. (2016). Joint multimodal learning with deep generative models. *arXiv preprint arXiv:1611.01891*.

[6] Saeidnia, H. R. (2023). Welcome to the Gemini era: Google DeepMind and the information industry. *Library Hi Tech News*.

[7] Microsoft. Copilot. https://copilot.microsoft.com/.

[8] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122-1136.
[9] Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350-361.
[10] Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277-304.
[11] Lawlor, P., & Chang, J. (2023). The generative AI economy: Worth up to $7.9T, Qualcomm. https://www.qualcomm.com/news/onq/2023/11/the-generative-ai-economy-is-worth-up-to-7-trillion-dollars.
[12] Chui, M., Yee, L., Hall, B., & Singla, A. (2023). The state of AI in 2023: Generative AI's breakout year. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.
[13] Manduchi, L., Pandey, K., Bamler, R., Cotterell, R., Däubener, S., Fellenz, S., ... & Fortuin, V. (2024). On the Challenges and Opportunities in Generative AI. *arXiv preprint arXiv:2403.00025*.
[14] Generative Artificial Intelligence: The impact on intellectual property crimes. (2023, November 30). Eurojust.
[15] Lucchi, N. (2023). ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 1-23.
[16] McIntosh, D. (2018). We need to talk about data: how digital monopolies arise and why they have power and influence. *J. Tech. L. & Pol'y*, *23*, 185.
[17] European Commission. Data Act. https://digital-strategy.ec.europa.eu/en/policies/data-act.
[18] European Commission. European Data Governance Act. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act.
[19] European Commission. The Digital Markets Act: ensuring fair and open digital markets. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.
[20] European Commission. The Digital Services Act. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.
[21] European Commission. Artificial Intelligence Act. https://artificialintelligenceact.eu/the-act/.
[22] European Commission. Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[23] Ashmore, R., Calinescu, R., & Paterson, C. (2021). Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Computing Surveys (CSUR)*, *54*(5), 1-39.
[24] Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, *15*(8), 260.
[25] Bond-Taylor, S., Leach, A., Long, Y., & Willcocks, C. G. (2021). Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE transactions on pattern analysis and machine intelligence*, *44*(11), 7327-7347.
[26] Van Den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016, June). Pixel recurrent neural networks. In *International conference on machine learning* (pp. 1747-1756). PMLR.
[27] Rethage, D., Pons, J., & Serra, X. (2018, April). A wavenet for speech denoising. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5069-5073). IEEE.
[28] Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., ... & Yang, M. H. (2023). Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, *56*(4), 1-39.
[29] Ashish, V. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*, I.

[30]Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.
[31]Vrbančič, G., & Podgorelec, V. (2020). Transfer learning with adaptive fine-tuning. *IEEE Access*, *8*, 196197-196211.
[32]Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.
[33]Wegner, P. (1996). Interoperability. *ACM Computing Surveys (CSUR)*, *28*(1), 285-287.
[34]Bourreau, M., & Krämer, J. (2022). Interoperability in digital markets: boon or bane for market contestability?. *Available at SSRN 4172255*.
[35]Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277-304.
[36]Hoffmann, J., & Otero, B. G. (2020). Demystifying the role of data interoperability in the access and sharing debate. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *11*, 252.
[37]Otto, B. (2022). The evolution of data spaces. In *Designing data spaces: The ecosystem approach to competitive advantage* (pp. 3-15). Cham: Springer International Publishing.
[38]ISO. ISO/IEC 19941:2017. https://www.iso.org/standard/66639.html.
[39]European Commission. The European Interoperability Framework in detail. https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail.
[40]Brown, I. (2020). Interoperability as a tool for competition regulation.
[41]Aly, M., Khomh, F., Guéhéneuc, Y. G., Washizaki, H., & Yacout, S. (2018). Is Fragmentation a Threat to the Success of the Internet of Things?. *IEEE Internet of Things Journal*, *6*(1), 472-487.
[42]Kerber, W., & Schweitzer, H. (2017). Interoperability in the digital economy. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *8*, 39.
[43]Lundqvist, B. (2017). Standardization for the digital economy: the issue of interoperability and access under competition law. *The Antitrust Bulletin*, *62*(4), 710-725.
[44]OECD. Algorithmic Competition. https://www.oecd.org/daf/competition/algorithmic-competition-2023.pdf.
[45]Motwani, S. R., Baranchuk, M., Strohmeier, M., Bolina, V., Torr, P. H., Hammond, L., & de Witt, C. S. (2024). Secret Collusion Among Generative AI Agents. *arXiv preprint arXiv:2402.07510*.
[46]Fish, S., Gonczarowski, Y. A., & Shorrer, R. I. (2024). Algorithmic Collusion by Large Language Models. *arXiv preprint arXiv:2404.00806*.
[47]Mulligan, C. E., & Godsiff, P. (2023). Datalism and Data Monopolies in the Era of AI: A Research Agenda. *arXiv preprint arXiv:2307.08049*.
[48]Federal Trade Commission. (2023). Generative AI Raises Competition Concerns. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns
[49]Lucchi, N. (2023). ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 1-23.
[50]Novelli, C., Casolari, F., Hacker, P., Spedicato, G., & Floridi, L. (2024). Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity. *arXiv preprint arXiv:2401.07348*.
[51]Carugati, C. (2024). The Generative AI Challenges for Competition Authorities. *Intereconomics*, *59*(1), 14-21.
[52]Federal Trade Commission. (2024). FTC launches inquiry into generative AI investments and partnerships. https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships.

[53]Soni, V. (2023). Impact of Generative AI on Small and Medium Enterprises' Revenue Growth: The Moderating Role of Human, Technological, and Market Factors. *Reviews of Contemporary Business Analytics*, *6*(1), 133-153.
[54]Fazekas, M., & Tóth, I. J. (2016). From corruption to state capture: A new analytical framework with empirical applications from Hungary. *Political Research Quarterly*, *69*(2), 320-334.
[55]Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, *65*, 277-301.
[56]Carugati, C. (2023). Competition in Generative AI Foundation Models.