International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03927
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A Review of Current Concerns and Mitigation Strategies on Generative AI and LLMs

Ruchika[1], Hemant Singh[2], Astitva Singh[3], Himanshu Bansal[4], Sanna Mehraj Kak[5*]

[1,2,3,4,5] Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India
ruchika@niet.co.in, hemant31400@gmail.com, astitvasingh4122@gmail.com, 0211csai061@niet.co.in, sanah.mehraj@gmail.com

Abstract

The upcoming of the large language models and generative artificial intelligence had Completely change the way in which we generate and understand language, and also start the beginning of a new phase in AI-driven applications. This review paper over see the advancements and changes that have occurred over time, providing a thorough assessment of generative artificial intelligence and large language models, while we also look upon their impactful potential across different areas. The first section of the research focuses on the changes of extensive language models and generative AI, and we will try to focus upon developments in models like GPT-4 and others. These models have shown their ability number of times from applications in various sectors, from automated content generation to acurate conversational agents.

They are characterized by their capability to produce text that is both coherent and contextually appropriate. However, despite their accuracy, strengths, generative artificial intelligence and large language models face critical ethical, technological, and societal issues. Some main stream concern arises from the biases present in the training data, which can cause and lead to social inequalities.Here we looks into the causes of these biases and their implications, stressing the need for comprehensive frameworks to identify and mitigate them.

Keywords: **backpropagation, bert, diffusion models, explainable ai (xai), generative ai, image synthesis, long short-term memory (lstm), natural language processing (nlp), neural network, recurrent neural network (rnn), small language model (sml),** and **transformer model.**

Introduction: The period characterized by advancements in artificial intelligence has given rise to influential large language models, including OpenAI's GPT-4, Google's Pathways Language Model (PaLM 2), Anthropic's Claude, and Meta's LLaMA 2, among others. These large language models have the potential to change how we create and interpret information across different fields. However, despite their significant capabilities, generative artificial intelligence and large language models also pose considerable ethical, technological, and societal challenges. A major concern is the biases found in the training data, which may exacerbate social injustices. This study investigates the origins of these biases and their repercussions, underscoring the importance of developing comprehensive frameworks for detecting and preventing bias. Additionally, the review addresses the privacy and security issues associated with the deployment of these models, examining potential vulnerabilities and strategies for mitigation.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03927
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

1.Introduction. In the era of development of artificial intelligence has lead emergence of powerfull large language models, such as **gpt-4** from **openai, pathways language model (palm** 2) from **google, claude** from **anthropic, large language model meta ai (llama** 2) from **meta** [1,2] etc. **Large** language models, are capable of transforming the way we generate and process information across various domains

1.1 Evolution and Key Milestones in Generative AI and LLMs:

The roots of generative AI can be traced back to fundamental concepts in neural networks and machine learning. The groundwork for more intricate models was laid by early AI research, which primarily focused on rule-based systems and fundamental statistical methods. A significant turning point was the advancement of neural networks, particularly the backpropagation algorithm, which emerged in the 1980s. Multi-layer neural networks were trained more efficiently using backpropagation, so that the models could identify complex patterns within the input data.

Models such as bert (bidirectional encoder representations from transformers) by devlin et al., and gpt-2 (generative pre-trained transformer 2) by openai, proved the effectiveness of llms with their release in 2018 and 2019 respectively[3,4]. A major advancement in LLMs was marked with the emergence of GPT-3 by OpenAI in 2020. With 175 billion parameters, gpt-3 could handle a wide range of tasks with very little tuning. With the release of GPT-3, Generative AI models like GPT-4 and others have been developed to check out what these technologies can achieve after improvements

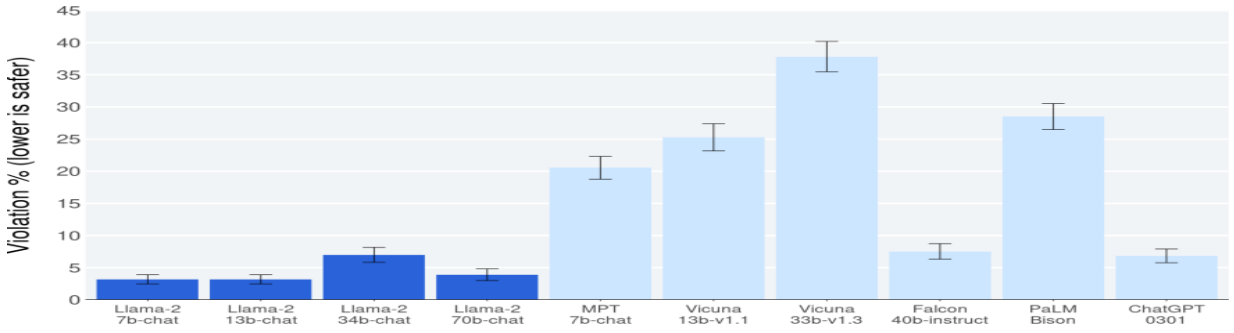

Figure 1: The image compares the number of violations and content-related outbreaks across various large language models (LLMs), including ChatGPT, Bard, and LLAMA.

1.2 Capabilities and Applications of Generative AI and LLMs:

Multiple industries are being transformed using various Generative AI tools and LLMs (Large Language Models). Some core capabilities of these tools and models include generating text in human-like manner, creating realistic visuals, generating audio and video content from text provided**.** GPT-4 and other models before it were able to generate meaningful and relevant text that set new standards in NLP (Natural Language Processing) [5 **].** Numerous applications are developed because of its ability to generate natural language that enhance user experience.

1.3 Architectural Innovations and Advancements:

With the creation of more AI-specialized tools and LLMs (Large Language Models), and advancements in their architecture there is also a impact on NLP (Natural Language Processing)**.** In 2017 Vaswani et al. introduced the transformer architecture. Transformers **utilize** a self-attention **mechanism** to process **entire** sequences of input **simultaneously, unlike traditional** recurrent neural networks **(RNNs)** [5,6]

Which processes data **in a sequential manner. This invention has made model** training more scalable and **efficient, revolutionizing the field.** The model **can** better **understand the connections between words in a sentence and their importance** by **using the self-attention mechanism, which weighs** the value of each **word. In** 2019, **openai introduced Gpt-2, a groundbreaking** unidirectional, autoregressive **model that demonstratedthe immense capabilities of** text **generation.** The **ability** of **gpt-2** to **produce** content that is both coherent and contextually **relevant attracted significant** attention **[7].** With the **introduction** of **gpt-4** in 2022, which **featured** even larger datasets and more advanced training methodologies, the **progress** in coherence, contextual comprehension, and application variety **was demonstrated** by **gpt-4.[8].** These architectural **advancements** have **expanded** the **scope** of generative **ai** and **llms, enabling their use in various** other **fields**.

1.4 Environmental Impact of Training Large-Scale Models

Large language models (LLMs) and generative AI had changed so quickly , but at a some related as the environmental cost. It takes a lot of computer power to train these models, we needed the constant thousands of powerful GPUs to operate nonstop for days or even weeks. Due to which extensive electricity use, this procedure has a significant carbon footprint.

A significant 2019 study revealed that the carbon dioxide emissions from training a single large-scale model, like BERT [8,9], can equal the lifetime emissions of five cars. The energy requirements of models have expanded exponentially from BERT to GPT-3 and GPT-4 as they have become larger and more complicated. The requirement to adjust these models for particular tasks increases the overall energy usage and worsens the environmental effect.

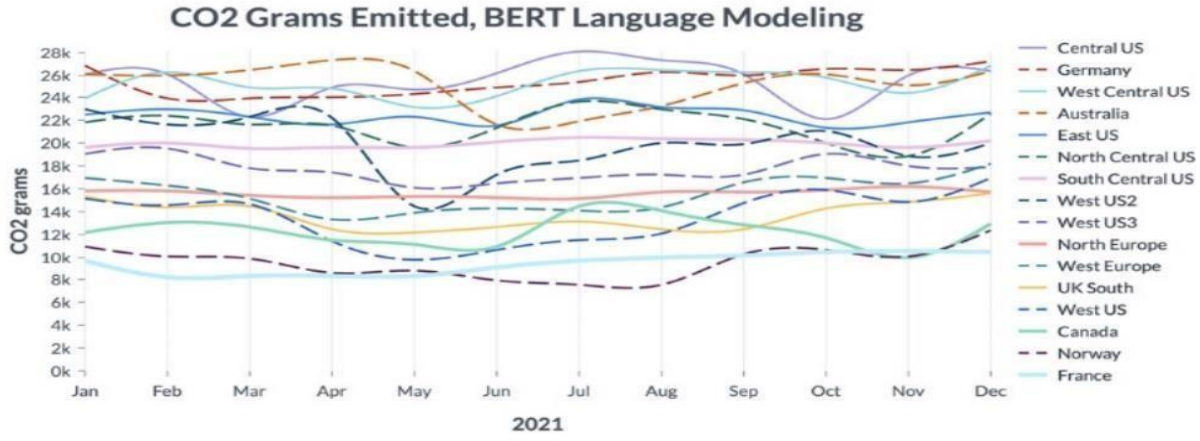

Figure 2: Carbon Footprint by various server of the large language model over the period on 2021

2..Literature Review

Significant development made in the field of Artificial Intelligence (AI) due to its quick development, especially in the areas of Generative AI and Large Language Models (LLMs). These models have completely changed how robots comprehend and produce writing that resembles that of humans. Examples of such designs are GPT-3 and GPT-4[10]. In literature study we will see the various aspects of generative AI and LLMs including the functions they perform, advancements in their architecture, their effects on environment, achievements, and applications in different fields.

Journey towards generative AI and LLMs started with the rule-based systems and statistical methodologies [12]. Neural networks that were introduced in 1980s came up with more advanced generative AI models [12-13]. With

the introduction of transformers by Vaswani et al. in 2017, it was possible to create effective and efficient LLMs. In 2019, OpenAI came up with GPT-2, and showcased its ability to generate meaningful and human-like text, which laid the groundwork for GPT-3 that was released in 2020. GPT-3 was tuned with 175 billion parameters and was capable of performing tasks like language translation and content creation. With further advancements, GPT-4 was released in 2023 and was able to understand images along with text. Today, these LLMs and generative AI tools can be seen in a wide range of applications.

They help produce articles, reports, and creative writing in content development, greatly cutting down on the time and effortneeded.

They are also used in code creation, where they aid developers in debugging and provide code snippets. By compiling voluminous medical literature and even helping with diagnosis, LLMs in the healthcare industry support medical research.

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$\text{Transformer}_{base}$	P100x8	1415.78	12	27	26	$41–$140
Transformer_{big}	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
BERT_{base}	V100x64	12,041.51	79	1507	1438	$3751–$12,571
BERT_{base}	TPUv2x16	—	96	—	—	$2074–$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	$942,973–$3,201,722
NAS	TPUv2x1	—	32,623	—	—	$44,055–$146,848
GPT-2	TPUv3x32	—	168	—	—	$12,902–$43,008

Figure 3: The Computational relation between Carbon footprint and the power consume at same time to different model to learn with that of cloud cost

LLMs and generative AI have shown impressive potential in a range of applications. These models are quite goodat comprehending natural language, which makes it possible to create sophisticated virtual assistants and conversational agents that can converse with people just like them.

Table 1: Quantitative Comparison of Generative AI Models and LLMs

Model	**Year of Release**	**Number of Parameters**	**Key Applications**	**Ethical Concerns**	**Environmental Impact (CO2 Emissions)**
GPT-2	2019	1.5 billion	Text generation, summarization, translation	Bias in generated text, lack of factual accuracy	Moderate (Training required large computational power)
GPT-3	2020	175 billion	Content creation, coding, customer service	Reinforcement of social bias, potential misinformation	High (~552 metric tons of CO2)

GPT-4	2023	Estimated 1 trillion	Multimodal AI (text, images), content generation	Privacy issues, bias in text and image processing	Very High (Larger dataset, increased carbon footprint)
BERT	2018	340 million	Language understanding, sentiment analysis	Gender and racial biases in contextual understanding	Moderate (Carbon emissions due to pre-training)
PaLM 2	2023	540 billion	Multilingual tasks, question answering	Reinforces existing biases in data, security concerns	High (Large computational resources for training)
Claude	2022	12 billion	Conversational AI, safety-focused applications	Limited interpretability, ethical safety limitations	Moderate (Focused on energy efficiency)
LLaMA 2	2023	70 billion	Research, open-source AI, High-quality image generation	Biased outputs, potential misuse in	Lower than GPT-3 (Designed for efficiency)

By compiling voluminous medical literature and even helping with diagnosis, LLMs in the healthcare industry support medical research [12-14]. These models also help in education sector by solving queries and providing real-time solutions. With the advancements in architecture,

LLMs and generative AI tools have shown significant advancement. With the introduction of the Transformer architecture, conventional architectures like LSTM (Long Short-Term Memory) and RNN (Recurrent Neural Network) were overtaken.

Transformers use self-attention technique to process input data in parallel that results in the increased efficiency and scalability. This technique set the foundation for creation of LLMs like GPT-2, GPT-3 and GPT-4. Using this technique GPT-3 that was tuned with 175 billion parameters [10] was able to generate meaningful and human-like text. Use of GPT's increased in a variety of fields when GPT-4 came up with multimodal features, enabling the modal to generate visuals along with the text.

LLMs and Generative AI have a exciting yet challenging future. One challenge is the development of model that can support multiple languages rather than just English that most of current models focus on. Researchers are also finding ways to make the decision-making process of these models transparent and making the model more understandable

2.1 Research Design

This section highlights advancements in LLMs across different sectors like healthcare, finance, and entertainment [3].It also highlights ethical, technological, societal challenges presented by these technologies. It is also important to have a discussion on responsible use of artificial intelligence as one can breach privacy and can generate harmful content. It is important in the advancement of these technologies as it links opinions of different people. For

example, European Union's AI act highlights strong ethical standards and AI systems accountability [14,15] also, the Pew Research Centre that concerned of deepfakes, highlights transparency and trust building [4]. It helps scholars and researchers to find solutions by minimizing risks and maximizing the use of generative AI as ethical considerations and regulatory awareness marks the basis for research [17-19].

2.2 Selection Criteria

The review carefully work upon the procedure and establish inclusion and exclusion criteria.And the peer review journal publication , conference , industry documented reports white research paper that already have been in market are used while making this paper the data belong from range of 2014 to 2024.

Table 2: Exploring Key Concerns in Generative AI and LLMs

Aspect	**Key Findings**	**Challenges/Issues Identified**	**Mitigation Strategies**
Model Architecture	Significant improvements in transformer-based architectures, e.g.,GPT, BERT, and LLaMA.	Scalability and Complexity increase resource demands.	Use of optimized training methods and parallel processing techniques.
Capabilities and Applications	LLMs excel in text generation, content creation, healthcare, coding, and education.	Bias in generated content, limited real-world understanding.	Implementation of Explainable AI (XAI), debiasing techniques, and improvement in multimodal capabilities.
Environmental Impact	High computational power for training LLMs leads to significant carbonemissions.	Large energy consumption during model training, contributing to environmental degradation.	Focus on green AI practices, use of energy-efficient hardware, and optimization techniques to reduce carbon footprints.
Ethical Considerations	Biases in data reinforce societal inequalities, e.g., gender or racial biasesin LLM outputs.	Propagation of harmful stereotypes and privacy concerns.	Use of diverse datasets, stricter frameworks for fairness and accountability, and continuous auditing for bias detection.
Privacy and Security	Models like GPT-3 and GPT-4 have raised concerns over data privacy and misuse,e.g., deepfakes.	Privacy breaches, generation of malicious or harmful content.	Development of better encryption, robust privacy frameworks, and regulations such as the EU's AI Act to ensure ethical AI

			usage.
Interpretability and Transparency	Difficulty in Explaining the model output nature.	Lack of transparency and accountability in AI decision-making.	Enhancing interpretability through post-hoc explainability methods, fostering transparent AI models that allow for human-understandable reasoning processes.
Technological andSocietal Impact	LLMs have transformed industries, including healthcare and contentcreation.	Displacement of jobs, misinformation, and a growing trust deficit in AI systems.	Public education campaigns, legislative interventions, and workforce reskilling programs to address AI-induced societal changes.
Future Research Direction	Multimodal and multilingual models are the future, aiming For broader applications.	Current models lack robustness in handling multiple languages and diverse domains effectively.	Focus on developing Multilingual LLMs, Expanding model adaptability, and improving global inclusivity of AI.

2.3 Ethical Considerations

Well with the upcoming of the generative AI and large language models (LLMs), the moral issues are essential while reviews of these technologies specific to those area which directly impact the mass people [19]-[20]. When ever we see the Large-scale datasets that may represent societal biases related to gender,racism, or socioeconomic status are frequently used to train generative AI and LLMs we need to be more careful in that sector.

This involves makingcertain that proprietary data is handled in accordance with ethical standards and that sensitive informationis anonymized. Maintaining the privacy of the data sources and the participants in the research upholds the legitimacy of the evaluation and safeguards the rights of the people whose information is disclosed.

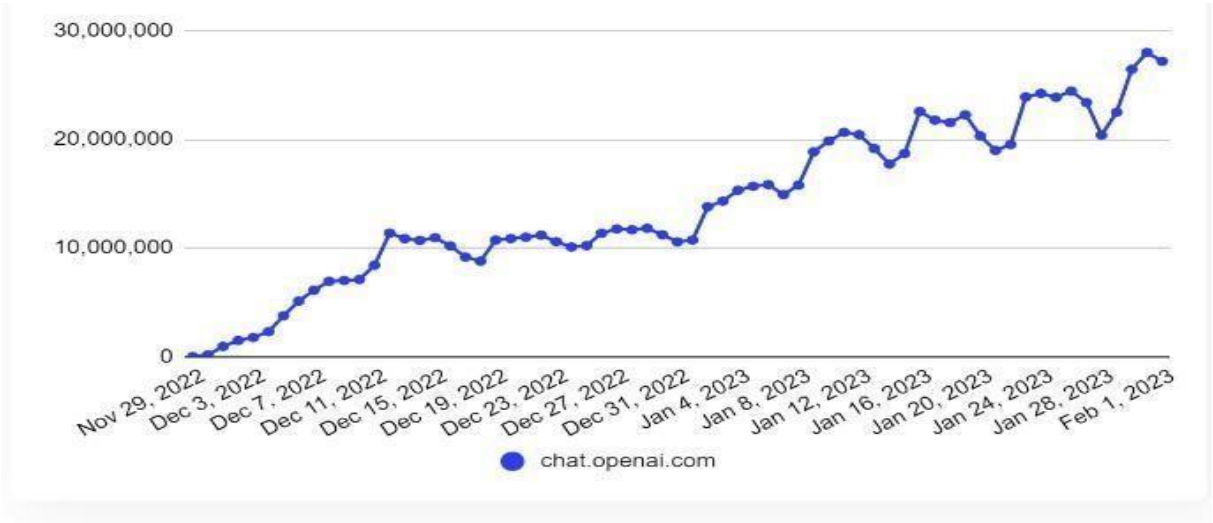

Figure 4: Represent the increase in model data the number of false outputalso havesurge due to poor data quality

Finally, although though it's usually not necessary to get informed consent from the authors of the reviewed literature, it's important to confirm that any research involving human beings was carried out under the proper ethical supervision. This guarantees that thereview upholds the highest ethical standardsand that the rights and welfare of the original study participants are respected[14].

The review hopes to respect the highest standards of research integrity by abiding by these ethical considerations and responsibly advancing generative AI and LLMs in a way that is just, open, and advantageous to society.

3.Result and Discussion

The findings highlight key ethical dilemmas, including privacy risks, the possibility of malicious use, and difficulties in maintaining accountability and openness. On the technical side, challenges like biased datasets, unclear decision-making processes in AI models, and inconsistent results further complicate the landscape. Societally, the rise of AI threatens employment opportunities, fuels the spread of false information, and erodes confidence in these systems.

To counter biases, the study suggests adopting strict ethical standards, increasing transparency, and using technical fixes such as refined bias-correction methods and clearer model explanations. For broader societal impact, proposals include public education initiatives, stronger legal frameworks, and support for workers adapting to AI-driven changes. While some solutions show potential, others expose unresolved issues in both theory and real-world application.

A deeper look at these strategies reveals their strengths, weaknesses, and contradictions, pointing to areas needing more study—particularly novel approaches and cutting-edge tech that could push past existing barriers. The discussion offers actionable insights for practitioners, lawmakers, and developers, stressing the need for ethical awareness in AI design and deployment.

In closing, generative AI and large language models (LLMs) hold immense promise across fields, yet their evolution—from GPT-2 to GPT-4—also raises pressing ethical, technical, and social questions. Issues like dataset biases risk reinforcing inequality, demanding robust safeguards to prevent discriminatory outcomes or

misinformation. Environmental costs also can't be ignored, urging greener AI training methods.

Ethical guardrails—transparency, accountability, and data privacy—must remain central. Policies like the EU's AI Act reflect growing consensus on responsible innovation. Moving forward, research should prioritize explainable models, multilingual support, and inclusive outputs. Collaboration among experts, developers, and regulators will be vital to harness AI's benefits while minimizing harm. Ultimately, a measured approach—one that balances progress with ethics and sustainability—will determine whether these technologies achieve widespread trust and long-term success.

4. Future Scope and Research Direction

As we continue to explore the boundaries of these technologies, several key research directions and opportunities emerge. One critical focus is enhancing the sustainability and efficiency of AI models. Researchers are actively working on energy-efficient hardware and algorithms to reduce the massive environmental footprint of training large-scale systems. Techniques like model pruning, quantization, and specialized hardware (such as GPUs and TPUs) play a vital role in this effort.

Another promising area is real-time adaptation and continuous learning. Future AI systems may evolve dynamically, adjusting to new data and user interactions without needing full retraining. This could lead to more personalized and responsive applications, powered by algorithms that learn incrementally.

Finally, innovative training methods—such as federated and self-supervised learning—could redefine how AI models are built. These approaches leverage decentralized data and reduce dependency on labeled datasets, potentially making AI more scalable and adaptable across different fields.

References

[1] Gvirtz, A., and O. A. Acar (2024). GenAI Can Help Level the Playing Field for Small Businesses. Harvard Business Review

[2] TechPark for AI (2023). Generative AI: Democratizing Creativity for SMEs & SMBs.

[3] Shen, K. N., Dwivedi, Y. K., D'Ambra, J., Sajib, S., Michael, K., Akter, S., and McCarthy, G. (2021). Algorithmic bias in AI-era data-driven innovation

[4] Jelstad Løvaas, B., Askeland, H., Espedal, G., & Sirris, S. (2020). Comprehending Leadership and Values in Organizations. Understanding Values Work, Askeland, H., Espedal, G., Jelstad Løvaas, B., & Sirris, S. (Eds). Cham, Palgrave Macmillan

[5] Belitski, M., and Audretsch, D. B. (2020). R&D and knowledge spillovers' contributions to productivity and innovation. Article 103391, European Economic Review, 123.

[6] Von Krogh, G., Hinds, P. J., Bailey, D. E., Faraj, S., & Leonardi, P. M. (2022). We Are All Technology Theorists Now: A Relational View of Developing Technology and Planning

[7] Content Team Bluebox (2024). Seven Crucial Risk Management Techniques All Singaporean SMEs Must Use

Bughin, J. (2023) [8]. Using AI: What Sets Superstar Companies Apart from Ordinary Businesses?

[9] Clear, F., and A. I. Canhoto (2020). A paradigm for identifying the potential for value loss using artificial intelligence and machine learning as business tools

[10] Tsao, H.-Y., Ferraro, C., Campbell, C., Sands, S., & Mavrommatis, A. (2020). From data to action: How Business Horizons and AI may be used by marketers

[11] Miremadi, M., Chui, M., & Manyika, J. (2018). What AI can and cannot accomplish for your company at this time.

[12] Dawson, G. S., Chenok, D., & Desouza, K. C. (2020). creating, developing, and implementing systems that use artificial intelligence

[13] Massey, C., Simmons, J. P., & Dietvorst, B. J. (2015). People mistakenly avoid algorithms after witnessing their errors, a phenomenon known as algorithm aversion.

[14] Demsar, V., Sands, S., Ferraro, C., Restrepo, M., & Campbell, C. (2024). The Conundrums of Customer Service Powered by Generative AI

[15] Forbes, 2023. 19 Useful Ways Generative AI Can Help Small Businesses.

[16] Saleh, T., Fountaine, T., & Mccarthy, B. (2019). Creating the AI-Powered Enterprise.

[17] B. Gates (2023). The Age of AI has begun: AI is just as revolutionary as the Internet and cell phones.

[18] Haenlein, M., and Kaplan, A. (2020). Come together, world rulers! Artificial intelligence's opportunities and difficulties

[19]. J. Kietzmann and A. Park (2024). Large Language Models, Conversational Chatbots, and Their Role in Business and Society, Written by ChatGPT

[20] Kuzlu, M., Gurler, N., Xiao, Z., Sarp, S., Catak, F. O., & Guler, O. (2023). generative artificial intelligence's ascent in the medical field.