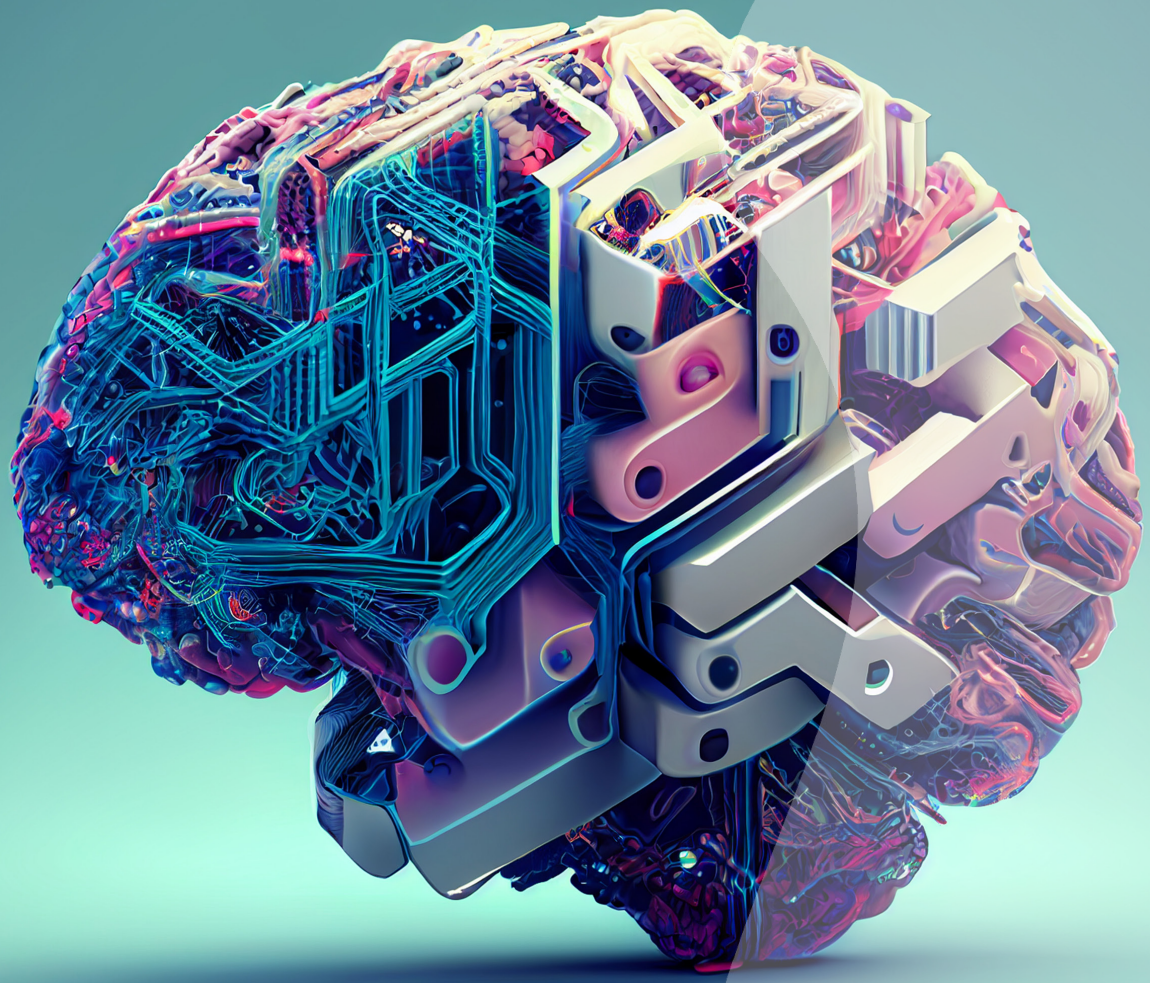

TechREPORT

Generative AI

September 2023

IDB

Authors:

Adrian Cevallos
Lucia Latorre
Gianfranco Alicandro
Zleste Wanner
Ignacio Cerrato
Jose Daniel Zarate
Juana Alvarez
Karla Villacreses
Maria Pfeifer
Mariana Gutierrez
Veronica Villanueva
Alberto Rivera-Fournier
Alex Riobo
Cristina Pombo
Fernando Puerto
Jorge Rodriguez

TechLab

The TechReports are an initiative of the Emerging Technologies Laboratory of the IDB's IT department, known as TechLab, which is in charge of exploring, experimenting, and disseminating information about new technologies to learn about their impact on the IDB Group and the LAC region.

Acknowledgments: The IDB team would like to thank all the individuals who participated in interviews and provided key information for this document.

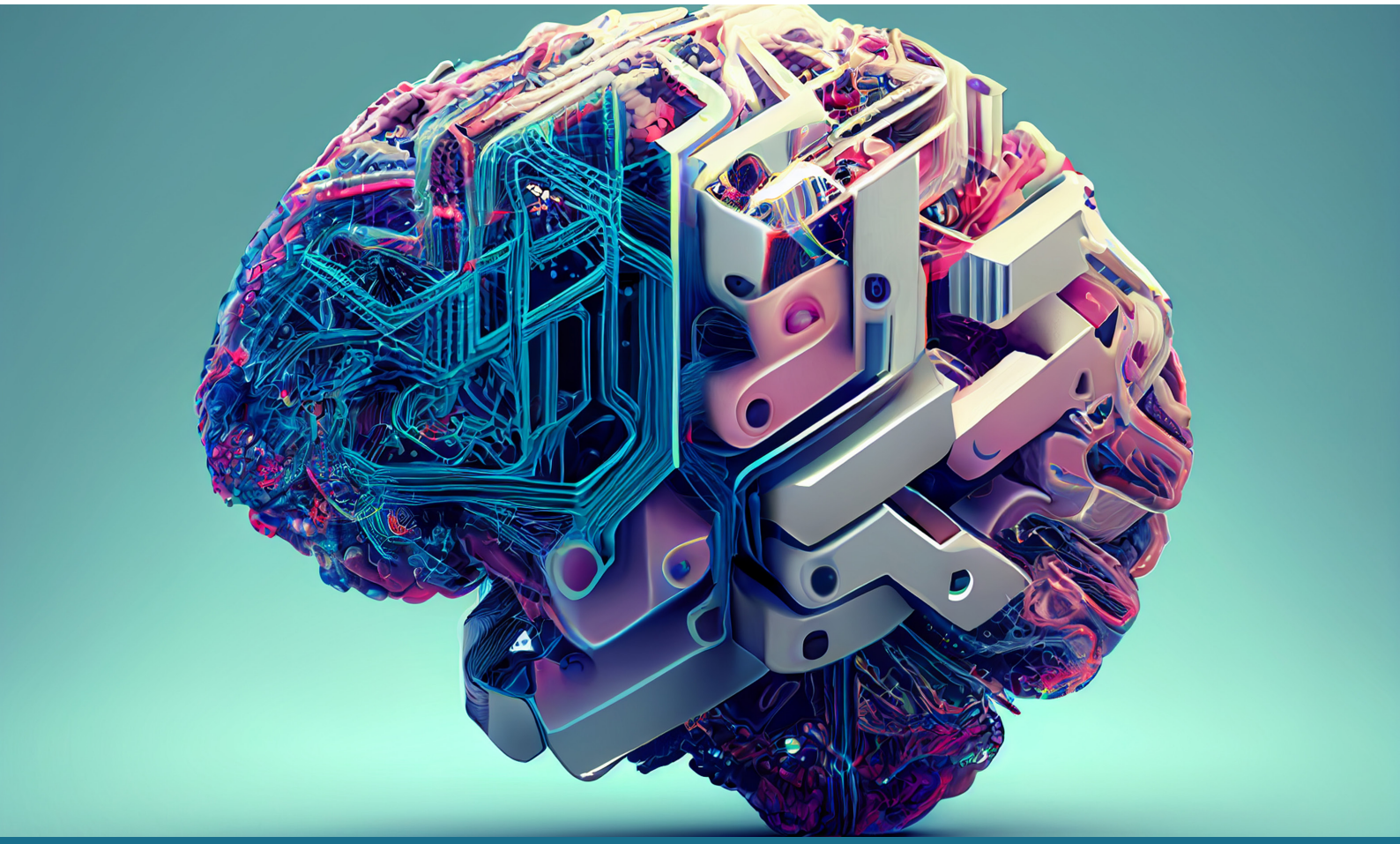

Table of contents

1. Executive summary

Generative AI is an emerging sub-domain of AI that is revolutionizing the use of technology as we know it. Its ability to generate new and unique content has great potential as a knowledge assistant, although it is still in the exploration phase.

Instead of simply classifying, analyzing, or processing existing data, Generative AI attempts to generate new data that resembles the original and is indistinguishable from that created by humans. To achieve this, Generative AI models use deep learning techniques, neural networks, and other advanced AI techniques to create models that can learn and replicate patterns in large datasets.

The IDB Information Technology Department, the Emerging Technology Lab, and the Ethics Department have analyzed both the use and risks of this technology and trends in the Latin America and Caribbean ecosystem.

AI generated image using the prompt:

"A stock image depicting ideas coming out of an isolated brain, orange background, no faces, no people"

2. Definition

Generative AI is a type of Artificial Intelligence (AI) that uses machine learning (ML) to produce new content from an extensive training dataset. The format of the result can be text, images, video, code, 3D renderings, or audio. Nowadays, when we interact with a search engine like Google or when we use a traditional question-answer chatbot, we are requesting existing information. In contrast, when using generative AI-based tools, the model is using existing information to generate original content, such as songs, poems, articles, etc.

Two flagship models that have put Generative AI in the general public conversation are ChatGPT and DALL-E. ChatGPT is a chatbot that can generate original text and DALL-E can create original images, both from a text input or "prompt."[1] According to Gartner, by 2025 generative AI will be responsible for producing 10% of the data, compared to less than 1% currently[1].

The organization behind these two models is OpenAI, a research and development company based in San Francisco, California. Due to the high popularity of ChatGPT, the company has already launched a paid subscription pilot called ChatGPT+[2] and closed a deal with Microsoft to license and commercialize the model within its suite of corporate products.

Other models similar to OpenAI's GPT include:

• Bloom: a text generation model in 46 languages and 13 programming languages with 176 trillion parameters, created by the BigScience project, a collective of more than 1,000 researchers from 60 countries.

• LLaMA, a "smaller" text generation model with between 7 and 65 trillion parameters, from MetaAI that requires less computational power than larger models.

The interesting thing about the technology behind Generative AI is that we cannot predict the content it generates, it is completely original and unique content, with reasoning that is becoming increasingly close to that of a human. Despite these advances in technology, and how useful it may seem for our day-to-day lives, we must remember that it is a kind of knowledge assistant. It assists us in generating content, but it cannot make decisions for us, nor should we base our decisions solely on what is generated by engines like ChatGPT. It should also be noted that not all generated content is correct, the technology used can provide inadequate responses.

---

[1] Prompt is the text input users use to interact with Generative AI tools
[2] https://www.technova-cpi.org/images/Documenti-pdf/Top%20Strategic%20Technology%20Trends%20for%202022_Gartner_31gen2022.pdf
[3] https://openai.com/blog/chatgpt-plus

3. Applications

A. Image generation: the model can generate a collection of original images based on a detailed description such as environment, subject, style, or location. Some available tools include OpenAI's DALL-E[4] and Stable Diffusion.[5] In another case of image generation, the Generative Adversarial Networks (GANs) method can convert a low-resolution image into a high-resolution image.[6] This application can be useful in the healthcare sector for patient diagnosis, as well as for security and surveillance purposes. For instance, this method is beneficial for creating top-notch versions of medical resources that are not feasible to store in high-resolution format due to cost constraints.[7] On the editing side, Google Pixel's Magic Eraser[8] feature uses generative AI to automatically remove unwanted photograph elements and fill in space.

B. Text generation: the model can generate original text based on a description. For example, an article, an essay, a script, a summary of a specific topic, etc. One of the most well-known examples is ChatGPT, where the model can hold a conversation and generate relevant content based on the context of the search. However, these models are still in development and present multiple challenges to generate reliable information to be used without prior verification.

C. Audio generation: the model can generate original audio based on text or even from another piece of audio. For example, it could create audio training based on notes for the education sector, as well as generate narration using the same voice as the reference audio. Lastly, the model can generate musical pieces based on a large set of data. However, this application will have to overcome copyright legislation to be able to use the training data of other musical artists. For instance, Google Duplex[9] is a virtual assistant in a spoken format capable of understanding the person one is talking to almost perfectly, answering practically as a human. It can support call centers.

D. Video generation: the model can detect time and space in videos to generate a new sequence. This model could be used to identify anomalies in security and surveillance videos as it can identify the probability of new sequences. Meta's Make-A-Video[10] and Runway Research's Gen-1[11] are two publicly available applications that show advances in technology.

---

[4] https://openai.com/product/dall-e-2
[5] https://stability.ai/blog/stable-diffusion-v2-release
[6] https://research.aimultiple.com/generative-ai-applications/
[7] https://research.aimultiple.com/generative-ai-applications/
[8] https://blog.google/products/photos/google-photos-magic-editor-pixel-io-2023/
[9] https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html
[10] https://makeavideo.studio
[11] https://research.runwayml.com/gen1

E. Synthetic data generation: the model can generate a type of data called "synthetic," which is artificially generated and not derived from direct observations in the real world. This application is particularly interesting as it preserves the privacy of the owners of the data used to train the model. This application could be used in the healthcare sector to generate and analyze data for disease research while preserving patient privacy.

F. 3D modeling generation: the model can generate 3D versions based on 2D images or text. With this method, a "digital twin" can be built in the metaverse as part of the creation of virtual worlds. Some applications could include training for the construction, manufacturing, or healthcare sectors, as well as city and physical product design.

AI generated image using the prompt:

"People in Latin America and the Caribbean using virtual reality technology"

4. Potential uses of large language models

An LLM, or "Large Language Model," is a Generative AI model designed to understand and generate human-like text based on the input it receives. These models are trained on vast amounts of textual data, enabling them to answer questions, compose content, assist in research, and perform a multitude of other text-based tasks with impressive accuracy. The most popular example of an LLM is ChatGPT. Here are four types of different purposes for LLM:

Synthesis: Summarization of large amounts of documents involves applying natural language to understand, extract and analyze the information. The large language models help to summarize legal documents and knowledge reports in fields as for as financial services, research, legal, or healthcare

Text generation: Involves various applications, including drafting mass emails, scheduling meetings[12], drafting job offers, knowledge papers, doctor's visit documents, drafting policy documents, or creating legal contract drafts. Generative AI can also be used to provide recommendations for farmers[13], support designers, and design travel guides for travelers[14]. These models can also be utilized as a chatbot for customer services as they are able to understand human-like text and generate human-like outputs.

Semantic Searches: Combining with prioritization and extraction capabilities, Generative AI can help in information analysis, such as analyzing price risk and database classification, among others. Additionally, Generative AI can be leveraged for fraud detection, by the identification of suspicious patterns and anomalies in text-based data, such as emails, invoices, or reviews.

Computer Code Generation It can be leveraged to generate computer code in different programming languages such as Python, JavaScript, Go, Perl, PHP, among others.[13]

[12] https://www.businessinsider.in/business/news/here-are-some-of-the-companies-that-are-using-chatgpt-salesforce-air-india-snapchat-in-the-list/articleshow/98904862.cms
[13] Articulo de gartners sobre agricultores
[14] https://www.forbes.com/sites/bernardmarr/2023/05/30/10-amazing-real-world-examples-of-how-companies-are-using-chatgpt-in-2023/?sh=51b06e631441
[15] https://openai.com/blog/openai-codex

5. Generative AI at the IDB

Generative AI has the potential to impact and disrupt the LAC region across different sectors and industries. The IDB Group has encouraged and is supporting the adoption of this technology in the region and the organization.

To better address the challenges and opportunities of this technology, an internal task force was created to align and coordinate Generative AI-related initiatives. This task force is mandated to generate knowledge, keep the organization and Region informed and updated, raise awareness about opportunities and risks, define the Generative AI agenda for the IDB Group, and identify safe tools to develop, implement, and use.

The following are featured articles, events, and publications that reflect on the use and potential effects of Generative AI in the LAC region:

- Global Challenge to Build Trust in the Age of Generative AI
- ChatGPT and education: opportunity, challenge or threat?
- The power of ChatGPT to develop innovative public policies
- ChatGPT and the future of work
- Chat GPT and the future of SMEs in Latin America and the Caribbean
- From the Printing Press to ChatGPT: The Challenge Does Not Lie in Technology but in Education Inequality
- Talk sessions on the Application of Artificial Intelligence in Online Education

6. Requirements and observations on the use of generative AI

It is very important always to have a human supervising the content generated by a tool like ChatGPT, this is a term we refer to as “human in the loop.” Generative AI can assist in content generation, but a human must always be involved, both in generating the prompt and in reviewing the content. Therefore, it is important to indicate in the prompt itself that if ChatGPT does not know the answer, it should say, “I don’t know.” If not, the tool will generate a new response that may be unreliable.

Use of prompts: One of the aspects that can determine the quality of the result, in addition to the quality of the model, is the design of prompts to use. This is where we can see the role of humans in the interaction of technology. The popularity of models like ChatGPT and DALL-E has inspired experts in creative industries to create and share collections of prompts for generating content relevant to their industries. Training content is also being created, such as “The Art of ChatGPT Prompting: A Guide to Crafting Clear and Effective Prompts”[16].

In this sense, the big difference in ChatGPT is that the model uses context to generate content. Starting with a prompt, ChatGPT can follow a conversation and respond according to the context.

Finally, as a reference, we share the case of the Association for the Advancement of Artificial Intelligence (AAAI), which establishes that generative models do not meet the criteria for articles published by AAAI and cannot be used as a citable reference. Attribution of authorship carries responsibility for the work, which cannot be effectively applied to AI systems[17].

[16] hhttps://fka.gumroad.com/l/art-of-chatgpt-prompting?a=705657043
[17] https://aaai.org/about-aaai/ethics-and-diversity/

7. Best practices for a safe and responsible use of generative AI tools

When using generative AI Tools, users should ensure their interaction is not only productive but also ethical and accurate. These guidelines are designed to provide responsible usage:

1. Validate and verify the accuracy of the generated content, we recommend to verify any source quoted by Artificial Intelligence. The primary purpose of Generative AI is to create new and original content, sometimes this content can be inaccurate or biased. As such it should not be used as a primary and single source of information since outputs could be unreliable or based on outdated information.

2. Do not use non-public or confidential information until your organization has identified and approved compliant tools. Some Generative AI public tools store and use the information you share to enrich the model, and it will become available to other users. We recommend to not share personal data and be mindful of potential privacy breaches that could arise through indirect identification of individuals.

3. Ask questions to the tool, when possible, about how conclusions were reached. Embrace ongoing learning and training to understand the capabilities and limitations of these tools and apply them responsibly in the work environment.

4. Be mindful of any biased data that can lead to inaccurate outcomes or discrimination. Make sure that the generated content you share is aligned with your organization's values.

5. Be aware of Intellectual Property Rights. These tools extract and generate information taking into account multiple sources, and the content generated might not be copyright free.

6. Disclose when content (text/image/video) has been created using Generative AI tools. There are some suggestions on how to cite ChatGPT and other Generative AI sources in the MLA, APA, and Chicago Manual Style citations[18].

[18] https://guides.nyu.edu/data/ai-citations

8. Ethical considerations

The ethical and responsible use of public tools such as ChatGPT must be treated with the utmost attention. First, we must be cautious in sharing confidential information that is not available to the public. This includes our responsibility to protect the privacy of the personal data we handle. Since AI models are trained with data, including information users share as part of prompts and interaction with the tool, we must be careful about what information we share and be aware that this data will become part of the AI model's training database.

Second, we must be aware of the risks related to the use of intellectual property. The algorithm has been trained through a large volume of information, and we must bear in mind that perhaps in this tool and others that may emerge in the future, the information used to train the model was not impartial or that such information is protected by intellectual property. As users we have a duty to respect and responsibly use others' intellectual property, which means recognizing the work of others and, when appropriate, obtaining the necessary permissions to avoid plagiarism or copyright infringement. It is also possible that the model takes information and misinterprets it, thus resulting in an answer that may promote discrimination. For instance, the model may promote discrimination based on race, gender, ethnic group, or religion .

In addition, users must be aware that Generative AI may provide erroneous or incorrect information. AI models are trained with data, and it is important to consider the reliability and trustworthiness of the data sources, in addition to noting that some of models were trained with information up to the year 2021, so they might not have updated information as of today. Therefore, it is important to exercise caution when relying on it. The use of unreliable data sources can lead the model to generate misinformation or make decisions that are not in line with ethical or legal considerations. In this regard, you can ask ChatGPT itself what data it used, and it returns the links to the sources, which facilitates verifying the information. However, if the source is not verified, it is best to disregard the provided data.

9. Security and privacy

Generative AI proposes significant opportunities however it also surfaces important privacy and security concerns.

From a security standpoint, it is important to consider the sensitivity classification of any data input in Generative AI platforms (i.e., public, confidential internal, restricted, or highly restricted). All data exchanged in the use of Generative AI is reviewed by AI experts to improve systems and to "train" the model and could eventually become part of the content provided to other users. There are currently no options to establish contractual agreements with OpenAI in the use of ChatGPT that would ensure the protection and confidentiality of organizations information.

It is also important to highlight the risk of methods that, under the pretext of new Generative AI platforms, install malicious software with the aim of stealing confidential information or committing fraud. Therefore, it is important to take precautions when acquiring, installing, or using new Generative AI applications. When it comes to privacy, Generative AI presents several areas of concern:

It is important to consider your own privacy in your use of public tools since the information you share when engaging with these tools is stored and can be used for other purposes. The personal information you share becomes part of the tool's database and can be potentially disclosed without further notice to you. It is also important to consider the possibility the system could be breached, or your personal data could become publicly available because of system bugs or leaks.

Consider the privacy of other individuals. It is possible that you could be providing information that could either directly or indirectly identify an individual when combined with other information already possessed by the tool and breaching their privacy without their knowledge or consent.

Third, it is important to consider privacy implications when exploring the development of applications of Generative AI. There is heavy debate around the challenge that Generative AI poses to fundamental privacy principles regarding transparency, access, use and rights. These concerns start with the way the data used to train models is obtained, primarily that it does not follow basic privacy principles or even, depending on the particular jurisdiction, is done in a way that violates privacy legislation. For example, some countries have issued bans of Generative AI tools and lawsuits have been raised declaring that data has been gathered without providing the proper notice, without a mechanism to ensure its accuracy, without proper consent controls for minors and without a legal basis for collecting this information.

There is consensus that information should be provided to individuals about how their data is collected and processed. The term "explainable AI" has been used to refer to the need to explain how the system makes decisions in an easy-to-understand manner as a way to uphold transparency and fairness principles of privacy. Finally, Generative AI has presented a challenge for the exercise of individual data subject rights and how to enable a path for a person to exercise their choice to remove their personal data, which may be very difficult, if not almost impossible, without compromising the usefulness of the model.

These are just some examples of potential privacy concerns that arise from the use of Generative AI. Adopting a privacy-by-design mindset is an important step towards keeping these concerns top of mind and incorporating privacy protection from the offset.

AI generated image using the prompt:

"AI security, 3d illustration"

10. Technical Considerations

Integrating Generative AI into enterprise applications holds tremendous potential from a technical perspective and there are some considerations that stay consistent regardless of the specific AI framework, technology, or programming language used.

Please find below relevant considerations when integrating this technology in your innitiatives:

- It is essential to comprehend the specific task at hand and select the appropriate model, such as GPT-3 and all its variants, GANs[19], BERT[20] or VAEs[21], aligning with data requirements.
- We recommend to optimize model training and consider ethical considerations to avoid biases and harmful content generation.
- Generative AI requires high energy power to run these models. Be aware of energy consumption and carbon footprint impact when indiscriminately using this technology[22].
- Implementing role-based authentication, gives you control access to sensitive data, ensuring that only authorized users can interact with and manipulate private information. This adds an extra layer of security to your application, reducing the risk of unauthorized access and data breaches.
- Prioritize user experience testing and implement robust security measures, such as encryption and secure communication channels, to protect user data and generated content.
- Comply with data privacy regulations and stay informed about advancements to leverage Generative AI responsibly and effectively across diverse applications.
- Take into account that the quality of the answers generated by the models, will depend on the quality of the data and knowledge base that is provided to leverage generative AI responsibly and effectively across diverse applications.

[19] https://www.coursera.org/specializations/generative-adversarial-networks-gans
[20] https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html
[21] https://synthesis.ai/2023/02/07/generative-ai-i-variational-autoencoders/
[22] https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure?in_source=embedded-checkout-banner

TechREPORT

Generative AI

September 2023