Revista Paraguaya de Educación a Distancia, FACEN-UNA, Vol. 5 (4) – 2024, pp. 21-28

DOSSIER – INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN

Artículo Original *https://doi.org/10.56152/reped2024-dossierIA2-art3*

Exploring Generative Artificial Intelligence

Explorando la Inteligencia Artificial Generativa

Jesús-Arnulfo Barradas-Palmeros
Universidad Veracruzana, México
https://orcid.org/0009-0008-1187-4981
E-mail: zs23000652@estudiantes.uv.mx

Efrén Mezura-Montes
Universidad Veracruzana, México
https://orcid.org/0000-0002-1565-5267
E-mail: emezura@uv.mx

Abstract

This work reviews the concept of Generative Artificial Intelligence (GAI), its main characteristics, and its classification inside the Artificial Intelligence (AI) field. In this way, the main GAI foundational algorithmic architectures for data generation are presented; each considers a previously gathered data set and generates novel and synthetic data. Subsequently, use cases for generative models reported in the literature are presented, ranging from applications in healthcare, education, and customer experiences. The deployment of GAI models still needs to overcome a series of challenges inspected in this document, considering the presence of bias, lack of transparency, and hallucinations presented by the models. Finally, the risks and misuse of GAI are reviewed since widespread use of this type of technology is expected in society.

Keywords: Artificial Intelligence, trend analysis, technology.

Resumen

Este trabajo presenta una revisión del concepto de Inteligencia Artificial Generativa (IAG) y su clasificación en el campo de la Inteligencia Artificial (AI) destacando sus principales características. Asimismo, se presentan las principales arquitecturas usadas por los algoritmos usados por la GAI para la generación de datos sintéticos y novedosos a partir de un conjunto de datos reales recopilado previamente. Posteriormente, se analizan algunos casos de uso presentados en la literatura respecto a modelos generativos en diferentes campos de estudio como lo es en la medicina, la educación y experiencias de cliente. El despliegue de modelos generativos requiere que se superen algunos desafíos como la presencia de polaridad, falta de transparencia y alucinaciones de los modelos, mismos que también son explorados en este documento. Finalmente, los riesgos y aplicaciones dañinas de la GAI son mencionados dado que es esperado un amplio uso de este tipo de tecnologías en la sociedad.

Palabras clave: Inteligencia Artificial, análisis de tendencia, tecnología.

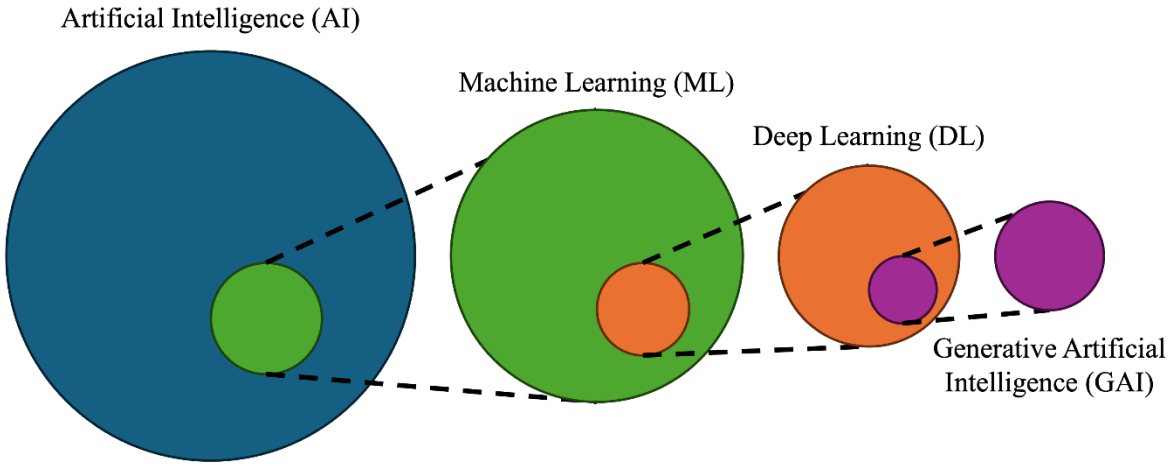
In recent years, it has been possible to see the rise of Artificial Intelligence (AI) based applications capable of generating impressive results in content creation, giving place to an AI field called Generative Artificial Intelligence (GAI) and attracting considerable interest. Large Language Models (LLMs) and text-to-image generation demonstrate their potential to produce synthetic content that is not distinguishable from human-generated content (Ooi et al., 2023).

Recibido: 07/03/2024 *Aceptado: 26/09/2024*

The previous has raised some opportunities, excitement among society, and a series of concerns about the future of this kind of technology (Gupta et al., 2024).

Machine Learning is a field of AI, as Russell and Norvig (2021) stated, where the computer builds a model taking some data as a basis, i.e., the computer learns from the provided data. The model is then used to solve a problem. There is a variety of algorithmic designs for the model that the computer uses. One of the approaches is to simulate, in a simple way, the inner workings of the brain, creating Artificial Neural Networks (ANN). When the number of layers of artificial neurons forming the ANN increases, the model is classified as a Deep Learning (DL) model (Moore et al., 2021). The most common architectures for GAI are formed with complex ANN architectures from DL, as will be seen in future sections of this document. Figure 1 presents a schematic of the concepts and fields of AI presented in this paragraph to shed light on the terminology and classifications used to refer to GAI.

Figure 1. From Artificial Intelligence to Generative Artificial Intelligence



As seen in Banh and Strobel (2023), the primary tasks of ML, after model construction from data (learning), correspond to group elements (clustering), assign a label to an observation (classification), and estimate a numeric value for an instance (regression). Nonetheless, GAI performs a different task, generating content in the form of text, images, audio, video, or a combination of the previous. The thing that distinguishes GAI is the ability to produce novel creations synthesized after the model is trained with existing data and information (Hussain, 2023). Given the previous, there is even a debate about whether GAI can be considered creative, as presented in Haase and Hanel (2023).

A brief explanation of the foundation models for GAI will be presented next. Then, some use cases of GAI will be mentioned, with particular attention paid to the education domain, along with some challenges that GAI presents to be implemented and used in society. The potential benefits and use cases of GAI have also given place to its use in activities that take advantage of the technology and harm some people. More of that is discussed before this document's conclusions and final takes.

THE FOUNDATION OF GENERATIVE ARTIFICIAL INTELLIGENCE

There are different architectures for the models on which GAI is based (Alhabeeb & Al-Shargabi, 2024; Jovanović & Campbell, 2022). Some of the most common ones are briefly explained in this work, including the Generative Adversarial Networks (GANs), the Diffusion Models, and Transformers in the form of Generative pre-trained Transformers (GPT). All those models shared the characteristic of working in a probabilistic way. The previous presents the consequence of returning different results when the same instruction is given. The instructions for the generative models are often called prompts, and the user uses them to limit and detail the expected answer from the model (Banh & Strobel, 2023).

GANs were proposed by Goodfellow et al. (2014). Gui et al. (2023) state that GANs are formed with two models: a generator and a discriminator. The generator's role is to learn from the training examples and generate new ones, while the discriminator classifies examples as real or generated. The generator aims to improve the examples it generates to make them look like the real ones with enough quality to fool the discriminator and make it think that the generated example is like one of the real ones from the data. During training, adversarial learning is used to improve the capabilities of both the generator and the discriminator, making the former better at fooling and the latter at discerning artificially generated images (Aggarwal et al., 2021).

As presented by Jovanović and Campbell (2022) and Cao et al. (2024), Diffusion Models use two main processes: forward and reversal. The forward process adds noise to the data following a particular distribution. The reversal process gradually reconstructs the original data from the sample with noise. This way, data is generated from random noise using the reversal process to apply the learned denoising steps. Diffusion Models have proven their capabilities to achieve high performance in image generation.

The transformer architecture was proposed by Vaswani et al. (2017), and the main characteristic of this DL architecture is the use of mechanisms called attention. The advantage of the transformer architecture is the ability to consider dependencies within the data in a long range. The previous has become an essential part of the Large Language Models (LLMs), allowing the model to consider the influence and relation of words placed long before and after the word that is being analyzed. LLMs are based on a transformer architecture called GPT, where a transformer is pre-trained with a large amount of data without labels. GPT architectures work stochastically to predict the next word in a sentence that will most likely be placed given the previous words called tokens (Banh & Strobel, 2023).

USE CASES OF GENERATIVE ARTIFICIAL INTELLIGENCE

Given the potential shown by the GAI models, many use cases have been considered, and more are expected to be developed. As stated by Ivcevic and Grandinetti (2024), discussions have taken place regarding whether GAI will unfavorably affect human creativity or boost it and provide powerful new tools. This document presents some representative applications of the GAI models. The use cases are derived from GAI's capabilities in generating text, images, videos, and more.

GAI in healthcare is highlighted as a tool to automate tasks and improve doctor and patient interactions (Ooi et al., 2023). The GAI models present an opportunity to make it easier to look for specific information instead of going personally through a large amount of information, making a question to the model, and obtaining a fast answer. Gathering patient information before the interaction with the doctor is another opportunity. Nevertheless, any AI application in healthcare must be implemented cautiously, given the implications derived from sensitive implications such as patient health.

As seen in Hussain (2023), GAI applied to e-commerce presents some opportunities where it can facilitate and automate some parts of the work and offer a better experience for the client. Part of the previous is the generation of product descriptions and personalized marketing experiences for the customer, as well as generating product recommendations. An exciting and additional point is using GAI to virtually generate an image of the product's appearance as if the client were in a virtual try-on. For example, if a client is looking for a shirt, GAI can use an image to generate a previous view in a picture or an avatar of the client wearing the shirt.

The work by Liu et al. (2023) presents the potential of GAI in the specific domain of materials science. It highlights the potential of using GAI for data collection and data augmentation. However, data augmentation with the current state-of-the-art technology provided wrong data cases. An explanation of the previous is that the models lack domain knowledge, which needs to be improved. Another application is conducted to acquire information for an experiment. GAI can reduce the required time and provide the user with the information in the literature. These potential benefits for materials science are helpful in other science applications.

Gupta et al. (2024) present an interesting potential application of GAI, exploring the potential of molecular drug discovery. The models tailored for this use case aim to reduce the cost and time required while developing a new drug. The generative model will try to produce and explore new molecular models. Nonetheless, a drawback of this use case is the requirement for a large amount of data in the model training process. The generation of such datasets is an essential step in future work.

Other applications for GAI include personalized retail experiences for customers with user-specific advertisements and customer assistance (Hussain, 2023), image processing and content analysis (Gupta et al., 2024), 3D modeling of proteins (Jovanović & Campbell, 2022), and increasing workplace efficiency (Ooi et al., 2023). Another domain where GAI can significantly impact is education, which is explored in more detail in the next section of this document.

GENERATIVE ARTIFICIAL INTELLIGENCE IN EDUCATION

Given the potential shown by the GAI models, various applications in education have been explored. Personalized educational experiences are mentioned by Ooi et al. (2023). The previous is expected to improve the engagement and accessibility for all students and allow the educator to create class materials faster and easier. GAI models can work as assistants during classes in notetaking and summarizing the day's contents. GAI can also be seen as a tool for student productivity. An example of the previous is the assistance the GAI models can provide in writing (Shailendra et al., 2024). However, a nonethical use of the tools can result in a detrimental situation for the student knowledge acquisition process. A big challenge in these cases is in students' assessment, where academic misconduct can emerge (de Silva et al., 2024).

The inefficiency of detecting AI-generated content creates the need for an approach where the students are provided with instructions on the ethical use of AI tools. The concern of AI misuse may affect the teacher and learner relationship where trust is questioned. Two additional problems arise from using GAI in education: privacy and lack of empathy. The former concerns using the student's data as part of the training of an AI algorithm, especially when sensitive information is considered. On the other hand, using an AI model must deal with the model's inability to perceive the student's emotions, background, and context. The above established a debate where schools are considering whether to ban the use of GAI or embrace it and make it part of the new ordinary (Mao et al., 2024).

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) has issued a pair of competency frameworks that serve as a reference for implementing AI

technologies in education. The AI competency framework for students (Miao & Shiohira, 2024a) states the need to provide students with skills and knowledge to use AI-based tools and the values for an environment of co-creation with AI. In the framework, the adoption of AI by the students has three progression levels: to understand, to apply, and to create. In the AI competency framework for teachers (Miao & Shiohira, 2024b), it is expressed that the relationship between teacher and student has changed to a dynamic where AI also forms part. AI can help teachers with their tasks, enabling new teaching and learning forms and requiring them to perform new roles. Consequently, teachers must be supported when taking advantage of the capabilities of IA while dealing with the risks and misuse of AI in the classroom.

CHALLENGES IN THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE

The potential applications of GAI models present challenges to overcome for future uses of this type of technology, as presented by Banh and Strobel (2023). The GAI models can present biases affecting the decisions made with the generated content. The bias can be inherited from the training data or while algorithmically generating the model. If there are biases in the training data, the learning process will take after the biases. As Ooi et al. (2023) state, biases and prejudices can severely affect any sensitive application of GAI, such as healthcare and finance.

Another issue is the building of trust in the model. The model's credibility can be gained if there is an explanation of how the output derives from the inputs of the system, which will become interpretable (Liu et al., 2023). Jovanović and Campbell (2022) assert that the preprocessing of the data should be an essential step for the generative models to avoid the presence of stereotypes and other types of content that can negatively impact the output of the models. Additionally, the developers behind the GAI models should identify and proactively communicate possible risks and threats and the benefits of using the model.

Two more points to be considered are disclosed by Banh and Strobel (2023): the model hallucination and the societal impact of GAI. The former represents the errors a generative model can present in its output. These errors can seem like a coherent output but contain fundamentally wrong parts. For example, inaccurate arithmetic statements or wrong representations of parts of the body in a generated image. Further model refinement is expected before actual and unsupervised tasks. On the other hand, the societal implications that should be considered are the risks of misuse, the democratization of the model's accessibility, and the risk of job losses, given the changes provided by GAI.

RISKS AND MISUSES OF GENERATIVE ARTIFICIAL INTELLIGENCE

Ferrara (2024) states that the potential of GAI applications is enormous, but this also presents the risk of creating dangerous use cases for society. With the intention of deception, propaganda, or dishonesty, users can look to gain an advantage in various situations. Dishonesty can be shown, for example, in academic environments where students can create and solve their assignments using GAI models. Given this, the student learning process can be affected. Propaganda can be generated with politicians' fake content, for example, spreading misinformation in society. Deepfakes have been used in that situation, creating realistic synthetic content. Deception uses GAI for specialized scams and fake testimonies. Specialized scams where the voice of a relative is artificially generated asking for money are examples of misuse of GAI by scammers.

In the work from Patel et al. (2023), a deepfake is described as creating real-looking synthetic content, where GAI is used to superimpose the face and voice of a person in content where another person appears. This way, an action that a person never performed can be

generated in a video that looks like it was recorded in real life. As we have seen, GANs can generate synthetic content that is complicated to distinguish from actual content. Deepfake detection methods are being developed to avoid misuse, where a robust detector is expected. Nevertheless, developing better deep-fake detectors forms a cycle with deep-fake generators. Better generators make better detectors, which can be used to form better generators, similar to the training process of GANs.

One proposal to deal with the deep-fake problem is employing a mechanism to validate the actual origin of the content, as stated by Strickland (2024), using content credentials. The content credentials mechanism proposes using the content's metadata to verify its generation source. Nonetheless, this type of security mechanism requires all content sources to include the proposed metadata changes and GAI to include changes in the metadata to state that it was altered. Other proposals to overcome the misuse of GAI are presented by Ferrara (2024), including security mechanisms such as better authentication protocols, watermarks for GAI content, and source verification.

CONCLUSIONS

GAI presents itself as a tool capable of revolutionizing a series of daily life scenarios, automating some routinary tasks, and boosting the person's capabilities in others. It has been an impressive development of GAI models and the results that they can generate. Considering the previous and the fact that this technology is only expected to improve, the future looks exciting for future use cases of GAI in various areas. Better efficiency and time savings are positive things that can directly impact the workspace.

Nonetheless, there is a series of challenges to overcome with the deployment of GAI technologies, as presented in this document. Additionally, some valid concerns are derived from the potential misuse of the GAI technology. A time is starting when the human senses can no longer distinguish between real and synthetic contents, with the risks that this statement implies. There are going to be significant challenges ahead. In the same way, the labor of lawmakers is expected to regulate GAI and act when things are going in the wrong way for society.

Author Contributions: All authors contributed equally to the conception, development, and writing of the article.

REFERENCIAS

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, *1*(1), 100004. https://doi.org/10.1016/J.JJIMEI.2020.100004

Alhabeeb, S. K., & Al-Shargabi, A. A. (2024). Text-to-Image Synthesis With Generative Models: Methods, Datasets, Performance Metrics, Challenges, and Future Direction. *IEEE Access*, *12*, 24412–24427. https://doi.org/10.1109/ACCESS.2024.3365043

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*(1), 63. https://doi.org/10.1007/s12525-023-00680-1

Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P.-A., & Li, S. Z. (2024). A Survey on Generative Diffusion Models. *IEEE Transactions on Knowledge and Data Engineering*, 1–20. https://doi.org/10.1109/TKDE.2024.3361474

de Silva, D., Kaynak, O., El-Ayoubi, M., Mills, N., Alahakoon, D., & Manic, M. (2024). Opportunities and Challenges of Generative Artificial Intelligence: Research, Education, Industry Engagement, and Social Impact. *IEEE Industrial Electronics Magazine*, 2–17. https://doi.org/10.1109/MIE.2024.3382962

Ferrara, E. (2024). GenAI against humanity: nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*. https://doi.org/10.1007/s42001-024-00250-1

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2023). A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3313–3332. https://doi.org/10.1109/TKDE.2021.3130191

Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. *Data and Information Management*, 100066. https://doi.org/10.1016/J.DIM.2024.100066

Haase, J., & Hanel, P. H. P. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity*, *33*(3), 100066. https://doi.org/10.1016/J.YJOC.2023.100066

Hussain, M. (2023). When, Where, and Which?: Navigating the Intersection of Computer Vision and Generative AI for Strategic Business Integration. *IEEE Access*, *11*, 127202–127215. https://doi.org/10.1109/ACCESS.2023.3332468

Ivcevic, Z., & Grandinetti, M. (2024). Artificial intelligence as a tool for creativity. *Journal of Creativity*, *34*(2), 100079. https://doi.org/10.1016/j.yjoc.2024.100079

Jovanović, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, *55*(10), 107–112. https://doi.org/10.1109/MC.2022.3192720

Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., Li, M., Ma, S., Avdeev, M., & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *Journal of Materiomics*, *9*(4), 798–816. https://doi.org/10.1016/J.JMAT.2023.05.001

Mao, J., Chen, B., & Liu, J. C. (2024). Generative Artificial Intelligence in Education and Its Implications for Assessment. *TechTrends*, *68*(1), 58–66. https://doi.org/10.1007/s11528-023-00911-4

Miao, F., & Shiohira, K. (2024a). AI competency framework for students. In *AI competency framework for students*. UNESCO. https://doi.org/10.54675/jkjb9835

Miao, F., & Shiohira, K. (2024b). AI competency framework for teachers. In *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/zjte2084

Moore, S. K., Schneider, D., & Strickland, E. (2021). How Deep Learning Works: Inside the Neural Networks that Power Today's AI. *IEEE Spectrum*, *58*(10), 32–33. https://doi.org/10.1109/MSPEC.2021.9563965

Ooi, K.-B., Tan, G. W.-H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V.-H., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L.-W. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, 1–32. https://doi.org/10.1080/08874417.2023.2261010

Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., Aluvala, S., & Vimal, V. (2023). Deepfake Generation and Detection: Case Study and Challenges. *IEEE Access*, *11*, 143296–143323. https://doi.org/10.1109/ACCESS.2023.3342107

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach (Global Edition)* (4th Edition). Pearson.

Shailendra, S., Kadel, R., & Sharma, A. (2024). Framework for Adoption of Generative Artificial Intelligence (GenAI) in Education. *IEEE Transactions on Education*, *67*(5), 777–785. https://doi.org/10.1109/TE.2024.3432101

Strickland, E. (2024). This Election Year, Look for Content Credentials: Media organizations combat deepfakes and disinformation with digital manifests. *IEEE Spectrum*, *61*(01), 24–27. https://doi.org/10.1109/MSPEC.2024.10380467

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf