

# A Systematic Literature Review on Generative Algorithms and their Impact on Machine Learning

Flavio Amayo-Gamboa<sup>1,\*</sup>, Javier Gamboa-Cruzado<sup>2</sup>, Amanda Durán Carhuamaca<sup>3</sup>,  
Saul Arauco-Esquivel<sup>4</sup>, Edwin Mendoza Torres<sup>1</sup>, Magaly Aranguena Yllanes<sup>3</sup>

<sup>1</sup> Universidad Nacional de Trujillo,  
Escuela de Ingeniería Informática,  
Peru

<sup>2</sup> Universidad Nacional Mayor de San Marcos,  
Facultad de Ingeniería de Sistemas e Informática,  
Peru

<sup>3</sup> Universidad Nacional de Cañete,  
Facultad de Ingeniería,  
Peru

<sup>4</sup> Universidad Nacional Mayor de San Marcos,  
Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica,  
Peru

famayo@unitru.edu.pe, jgamboac@unmsm.edu.pe, aduran@undc.edu.pe,  
saraucoe@unmsm.edu.pe, emendoza@unitru.edu.pe, maranguena@undc.edu.pe

**Abstract.** This research analyzes the influence and development of Generative Algorithms within the field of Machine Learning (ML), a sphere that is garnering increasing academic and practical interest. The goal is to unveil the state of the art and discern the Generative Algorithms and their impact on ML. A systematic review methodology was adopted, assessing relevant studies published between 2017 and 2023, focusing on Generative Algorithms and their impact on ML. Through the consultation of databases such as Scopus, Web of Science, Science Direct, Springer Link, Google Scholar, and ACM Digital Library, and the application of exclusion criteria presented in the PRISMA Flow Diagram, 62 papers were selected and analyzed. The results highlight a marked presence of research in first-quartile journals, demonstrating the high quality and relevance of the topic. International collaboration emerges as a crucial pillar, with the United States and Canada leading in significant contributions. The most striking conclusions suggest a consolidation of Generative Algorithms as a prominent area of study, with projections towards their integration into novel fields such as quantum computing. The research concludes that global cooperation and institutional support are essential for the progress of ML, emphasizing the importance of adopting

collaborative and interdisciplinary approaches in future studies.

**Keywords.** Generative algorithms, machine learning (ML), deep learning, systematic literature review (SLR), generative models.

## 1 Introduction

The interplay between generative algorithms and Machine Learning (ML) constitutes a fascinating convergence of two fundamental fields within artificial intelligence. Despite their increasing relevance, there remains a significant knowledge gap regarding the impact and trajectory of generative algorithms in ML over time, underscoring the urgency for a systematic and detailed analysis of the influence of these algorithms in the ML landscape. The existing literature has tackled various dimensions of generative algorithms and their interaction with ML. Perdomo-Ortiz et al. [1] focus on unsupervised

generative models, highlighting their role in exploratory data analysis in the era of big data. Conversely, Fanfarillo [2] compares classical and quantum ML models, emphasizing the superiority of restricted Boltzmann machines and Born machines of quantum circuits in generative tasks. Moreover, Neo et al. [3] detail the use of ML in the design of photonic crystals. The research by Sanchez-Lengeling and Aspuru-Guzik [4] centers on generative molecular design, examining molecular representations and neural network architectures.

In the study by Bilodeau et al. [5], the authors propose an innovative method employing Variational Autoencoders for the creation of synthetic populations of micro-agents, surpassing previous techniques. Additionally, Fan et al. [6] analyze the susceptibility of deep neural network classifiers to be fooled by minimal alterations, proposing defensive strategies.

Similarly, Yao et al. [7] present RNA-GAN, which integrates gene expression profiles with generative models to produce more realistic tissue mosaics. Dahl and Sørensen [8] develop predictive ML models for price signals in financial markets, enhancing trading strategies through synthetic data. In the study by Carvajal-Patiño and Ramos-Pollán [9], the application of deep learning in generating medical images is reviewed.

The authors Paz et al. [10] propose a comprehensive model for the detection of textual misinformation on social networks. Lastly, Alhomayani and Mahoor [11] evaluate the efficacy of conditional generative models in improving classifiers for imbalanced datasets.

The realm of generative algorithms in Machine Learning (ML) constitutes an expanding area of research whose depth and scope have yet to be fully elucidated. Despite rapid advancements in the adoption of these technologies, a lack of clarity persists regarding their long-term effects and effective integration across various domains. The existence of knowledge gaps indicates limitations in the comprehensive understanding of their disruptive impact.

The aim of this review is to offer a critical and detailed insight into the field, spotlighting advancements, and challenges, and proposing strategic directions for future research. The study concludes with the presentation of conclusions and

suggestions for future work based on the findings obtained.

This paper is structured as follows: Chapter 2 presents the theoretical background; Chapter 3 describes the review methodology; Chapter 4 discusses the main results and findings; and Chapter 5 concludes with reflections and perspectives for upcoming research.

## 2 Theoretical Background

### 2.1 Generative Algorithms/ Generative Models

In the realm of artificial intelligence (AI), generative models have gained fundamental importance [67]. Their primary aim is to generate new data that faithfully emulates the input data. These models have experienced significant evolution since Turing's early contributions, progressing through Hidden Markov Models and Recurrent Neural Networks, to the contemporary Generative Adversarial Networks (GANs) [68]. GANs have been a milestone in the generation of synthetic data over the last decade, extending their impact beyond healthcare.

This technology exemplifies the ability of generative models to learn and sample from the implicit density functions of data. This advancement is pivotal for creating realistic data with practical value across various fields, including computational chemistry [4].

### 2.2 Machine Learning

The interrelation between generative models and Machine Learning is essential and highly interconnected. The study by Zhang et al. [67] highlights the crucial role of Alan Turing in the development of Machine Learning, a discipline focused on creating machines with the ability to learn and reason in a manner analogous to humans.

Turing's innovative vision has been instrumental in achieving significant advancements in automatic learning, including the development of algorithms capable of generating complex and emergent behaviors. This concept is a central pillar in the areas of neural networks and deep learning.

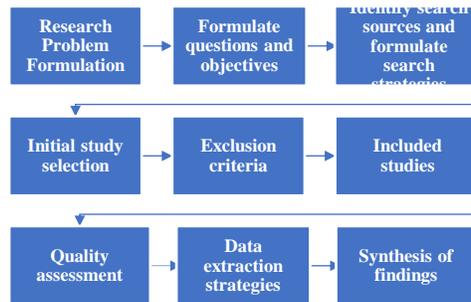


Fig. 1. Stages of the systematic literature review.

Table 1. Research questions and objectives

Research Question	Objective
<b>RQ1:</b> How are the quartiles distributed in the periodic publications that have addressed research on the Impact of Generative Algorithms in Machine Learning?	Determine the distribution of quartiles in periodic publications that have addressed research on the Impact of Generative Algorithms in Machine Learning.
<b>RQ2:</b> What terms (keywords) are predominant in the abstracts of studies related to the Impact of Generative Algorithms in Machine Learning?	Find the terms (keywords) that are predominant in the abstracts of studies related to the Impact of Generative Algorithms in Machine Learning.
<b>RQ3:</b> What sets of keywords tend to appear together in the literature examining the Impact of Generative Algorithms in Machine Learning?	Know the sets of keywords that tend to appear together in the literature examining the Impact of Generative Algorithms in Machine Learning.
<b>RQ4:</b> What categories of named entities, such as individuals, institutions, locations, dates, and figures, prevail in the abstracts of research on the Impact of Generative Algorithms in Machine Learning?	Determine the categories of named entities, such as individuals, institutions, locations, dates, and figures, that prevail in the abstracts of research on the Impact of Generative Algorithms in Machine Learning.
<b>RQ5:</b> What is the distribution of clusters of the conclusions of the papers distinguished by their high objectivity and minimal emotional polarity in the context of the Impact of Generative Algorithms on Machine Learning?	Understand the cluster distribution of the conclusions of papers that are distinguished by their high objectivity and minimal emotional polarity in the context of the Impact of Generative Algorithms in Machine Learning.

### 3 Review Method

In this systematic review, a methodology based on the guidelines by Kitchenham and Charters [29], recognized as a standard in academic research for the consolidation of scientific evidence, has been adopted. (See Figure 1).

#### 3.1 Research Problems and Objectives

For a detailed understanding of the objectives and scope of the research, Table 1 should be reviewed.

#### 3.2 Information Sources and Search Equations

Comprehensive searches were conducted in highly relevant academic databases, including

Scopus, Science Direct, Web of Science, ACM Digital Library, Springer Link, and Google Scholar. The selection of these databases was intentional due to their broad coverage in the fields of systems engineering, data science, and machine learning, key areas for the present review.

These databases provide access to high-quality, peer-reviewed studies and are recognized for their comprehensiveness in relevant academic literature.

Keywords such as generative algorithms, generative models, data science, data analysis, and machine learning were chosen for their importance in generative learning and its applications in data analysis.

Search equations were constructed to maximize the relevance of the studies and minimize the inclusion of irrelevant results. Inclusion criteria focused on high-impact studies,

**Table 2.** Information sources and search equations

Source	Search Equation
Scopus	TITLE-ABS-KEY ( ( "generative Algorithms" OR "generative models" OR "generative modeling" OR "generative Methods" ) AND ( "data science" OR "data analysis" OR "data mining" OR "machine Learning" ) )
Science Direct	Title, abstract, keywords: ("generative Algorithms" OR "generative models" OR "generative modeling" OR "generative Methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine Learning")
Web of Science	("generative Algorithms" OR "generative models" OR "generative modeling" OR "generative Methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine Learning") (Title) OR ("generative Algorithms" OR "generative models" OR "generative modeling" OR "generative Methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine Learning") (Abstract) OR ("generative Algorithms" OR "generative models" OR "generative modeling" OR "generative Methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine Learning") (Author Keywords)
ACM Digital Library	[[All: "data science"] OR [All: "data analysis"] OR [All: "data mining"]] OR [All: "machine learning"]] AND [[All: "generative algorithms"] OR [All: "generative models"] OR [All: "generative modeling"] OR [All: "generative methods"]]
Springer Link	('generative algorithms" OR "generative models" OR "generative modeling" OR "generative methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine learning")'
Google Scholar	("generative algorithms" OR "generative models" OR "generative modeling" OR "generative methods") AND ("data science" OR "data analysis" OR "data mining" OR "machine learning")

excluding those without empirical results or with limited access. Table 2 provides a comprehensive overview of the search strategies implemented.

### 3.3 Study Selection

The Paper Selection and Filtering process considered the following exclusion criteria (EC):

EC1 (Publication age): The publication dates were reviewed, excluding papers older than 7 years to ensure relevance in a rapidly evolving field.

EC2 (Language of the paper): Papers not written in English were eliminated, as English is the predominant language in international scientific literature.

EC3 (Type of study): Systematic reviews and bibliometric analyses were prioritized. Studies that did not provide a significant theoretical or empirical contribution were excluded.

EC4 (Full-text access): Papers whose full text was not available were excluded, as this would limit a thorough analysis of their content.

EC5 (Peer review): Studies published in journals and conferences with rigorous peer-review processes were selected.

EC6 (Clarity in titles and keywords): Papers whose title or keywords did not clearly reflect the content were eliminated to ensure alignment with the research topic.

EC7 (Abstract clarity): Papers with ambiguous or unclear abstracts were excluded, as the abstract must provide an accurate summary of the main contributions of the study.

EC8 (Originality): Studies that did not provide new perspectives or original data were eliminated, ensuring that the selected papers were relevant. Additionally, all duplicate papers were removed.

The selection process is detailed in Figure 2.

The InOrdinatio formula (1), derived from the document by Pagani et al. [42], was developed to assess the relevance of scientific papers:

$$InOrdinatio = \frac{H\_Index}{100} + (6 - Quartil) * 2 * [7 + (PublishYear - ActualYear)] + (\sum Ci), \quad (1)$$

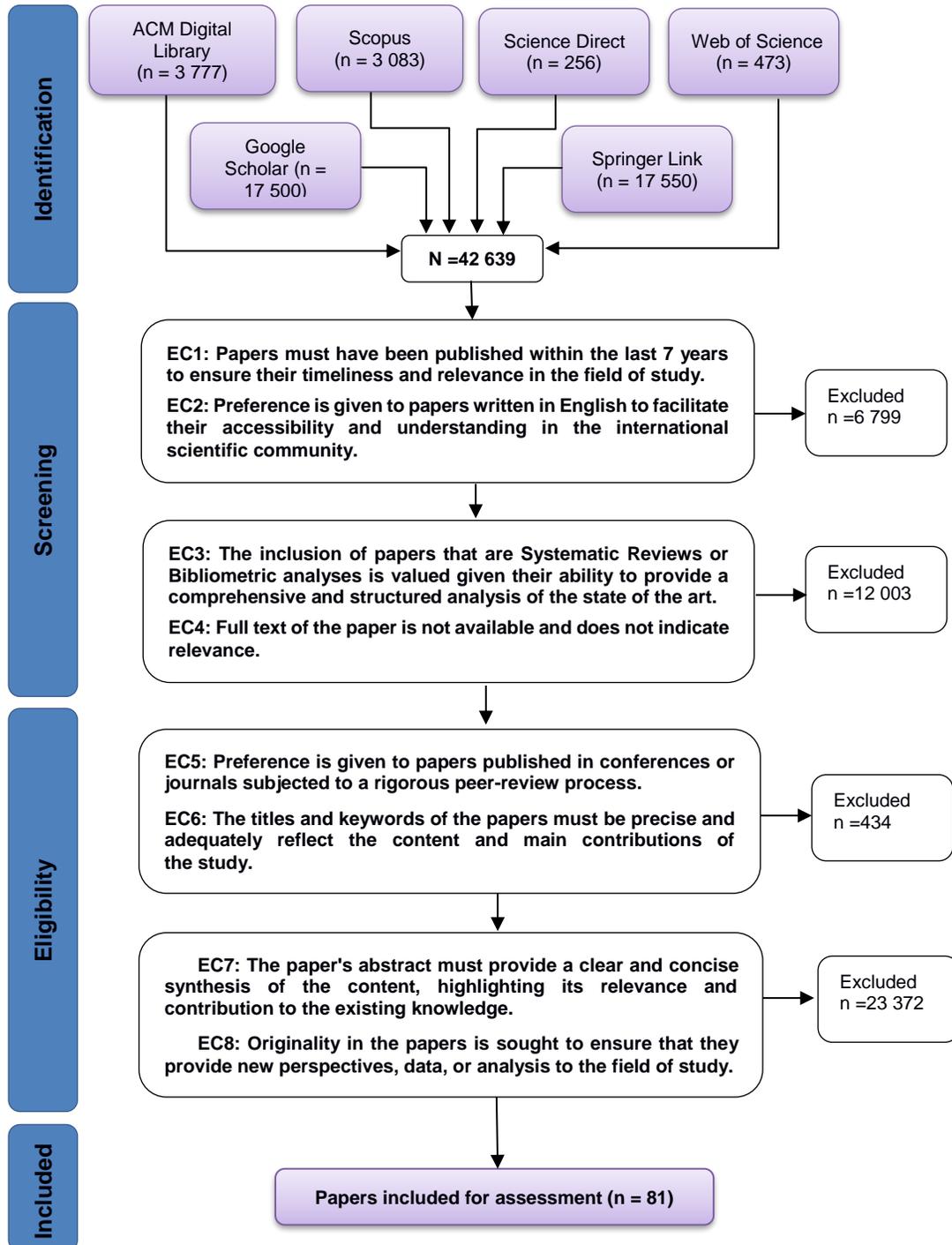


Fig. 2. PRISMA flow diagram

where:  $H\_Index$  is the paper's H-Index,  $Quartile$  is the paper's quartile, and  $\sum Ci$  is the total number of citations of the paper.

The index ranks the selected publications based on their relevance, combining four key factors: impact factor, year of publication, quartile, and citation count. By applying this formula, researchers can identify the most relevant studies within their bibliographic portfolio, ensuring a selection based on objective and meaningful criteria.

Studies that, despite their significant InOrdinatio score, were not available, were acquired by the researchers.

### 3.4 Quality Assessment

In the final stage of the research, specific criteria were applied to measure the quality (QA) of the studies, which was essential for determining the final selection of papers included in the analysis. Five quality criteria are detailed below:

- QA1. Does the paper present a logical and coherent structure that facilitates the understanding of the content?
- QA2. Are the data sets used in the research precisely specified?
- QA3. Are the results obtained in the research explicitly delineated and communicated?
- QA4. Do the study's conclusions align with the initial objectives?
- QA5. Does the study contribute valuable and applicable information for the advancement of the field of study?

In the quality assessment process of this study, 81 selected investigations were examined, following specific exclusion criteria. An assessment of each study determined that 62 studies met all the established quality criteria.

## 4 Results and Discussion

In this section of the study, the obtained results are presented and analyzed, placing them in the context of the previous literature and the objectives established in the research.

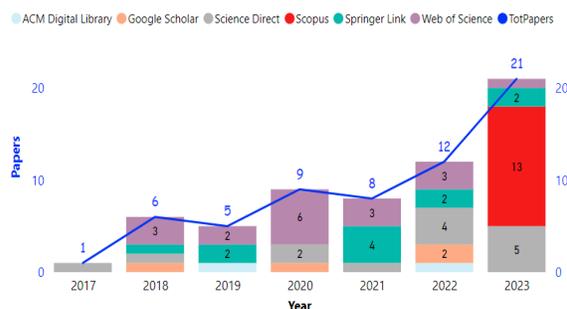


Fig. 3. Distribution of papers published by year

### 4.1 Overview of the Studies

The systematic literature review is a rigorous methodology that allows for the collection and evaluation of relevant studies on a specific topic. Once the relevant studies were collected, data extraction was performed to obtain detailed information about each study, such as the title, authors, publication year, and other relevant data. Figure 3 shows the number of studies published each year, providing a visual representation of the temporal distribution of research in this field.

This figure displays an increase in the number of publications on generative algorithms in Machine Learning from 2017 to 2023, highlighting the growing interest in this field. Scopus and Web of Science, which are consolidated as the main sources of these publications, reflect the preference of researchers or a greater inclusion of works of this nature in their indexes. The steady growth of publications between 2017 and 2023 can be attributed to several key factors.

Technological advancements, such as generative algorithms and machine learning, have gained popularity during this period, driving more research. Access to infrastructures like cloud computing and big data has enabled more complex studies, while increased funding for research in data science and machine learning has incentivized academic production. Additionally, changes in academic policies, such as the pressure to publish in indexed journals and the focus on open science, have contributed to this increase, allowing for broader international collaboration and access to resources that accelerate result generation.

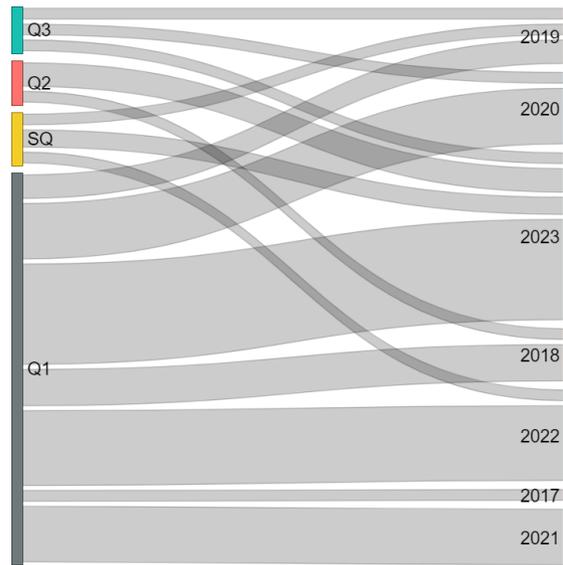


Fig. 4. Publications by quartile and by year

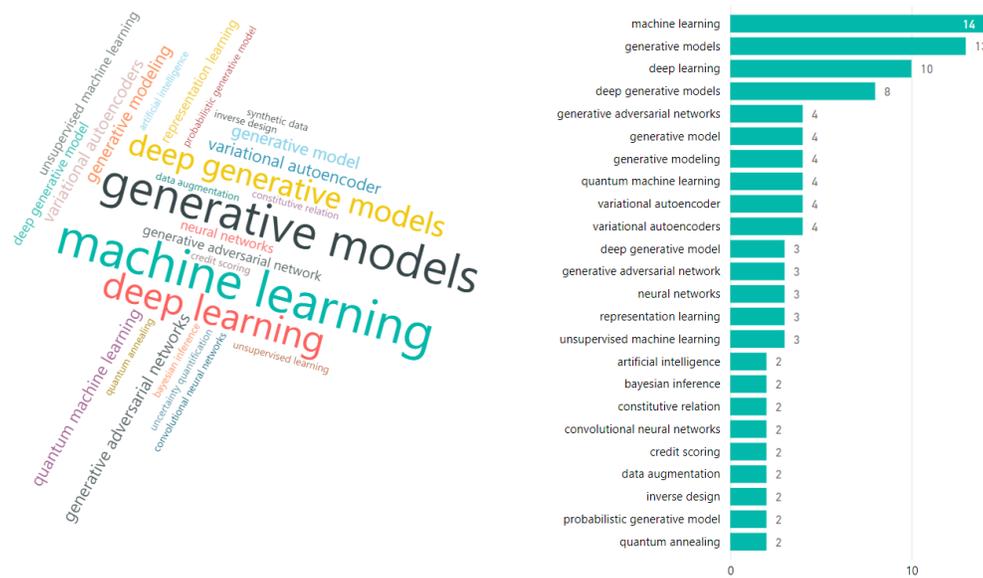


Fig. 5. Word cloud of the predominant keywords in the abstracts

When compared with the study by Cárdenas-Quispe et al. [10], which highlights IEEE Xplore as a primary source, a marked technical inclination in its collection is appreciated.

Conversely, Aparcana-Tasayco and Gamboa-Cruzado [5] show a pattern of sustained growth

until 2020 and a subsequent decrease, differentiating from the more uniformly ascending trend observed in this review.

In their study, Rojas Valdivia et al. [49] show similar behavior until 2021, followed by a drop in 2022, which could reflect changes in the dynamics

of publication or a shift in research focus towards emerging subdomains.

## 4.2 Responses to Research Questions

This phase of the Systematic Literature Review (SLR) is decisive, focused on exploring and detailing the responses to the four research questions (RQs) that have guided this study.

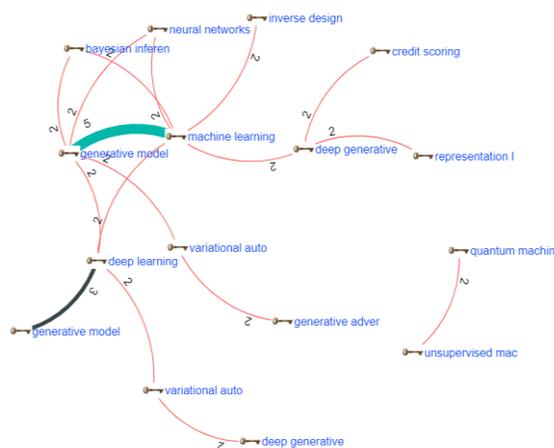
**RQ1:** How are the quartiles distributed in the periodic publications that have addressed the research on the Impact of Generative Algorithms in Machine Learning?

Quartiles (Q1, Q2, Q3, and Q4) are a classification that reflects the relative position of journals within their field, based on impact metrics. Journals in Q1 are those with the highest impact and prestige, while those in Q3 or Q4 have lower impact. SQ may represent publications outside the quartile system, such as conference papers or journals without a defined quartile ranking. Figure 4 displays the annual distribution of publications by quartile in the selected sources.

The temporal distribution of publications evidence trends in the research of generative algorithms and Machine Learning from 2017 to 2023. The persistence of papers in the third quartile (Q3) suggests established research in mid-range journals, while the increase in unclassified quartiles (SQ) and second quartile (Q2) in 2023 reflects an interest towards emerging areas and the growing perception of their relevance. Notably, the first quartile (Q1) dominates the distribution, particularly in 2023, underlining the advancement and increasing importance of this field.

The variability in quartiles observed in recent years could indicate increased competition in high-impact journals or a greater openness to explore publications in other impact levels. This behavior may also reflect the need to diversify publications to reach broader audiences or meet different evaluation metrics.

Publications in Q1 remain important, but the relevance of adjusting publication strategies to maximize reach and visibility across various academic forums becomes evident.



**Fig. 6.** Bibliometric network of keywords

**RQ2:** What terms (keywords) are predominant in the abstracts of studies related to the Impact of Generative Algorithms on Machine Learning?

The analysis of prominent keywords, presented in Figure 5, reveals the areas of greatest focus and their relevance in current research.

The analysis of the word cloud and bar chart shows that "machine learning" is the most predominant term, with 14 mentions, followed by "generative models" with 13 and "deep learning" with 10. This reflects the main focus of current research around these key concepts. The rise of terms like "deep generative models" (8 mentions) and "generative adversarial networks" (4 mentions) suggests that advanced data generation techniques are gaining relevance. Terms such as "quantum machine learning" and "probabilistic generative model," with only 2 mentions each, indicate that these topics are still in an emerging phase, though they are promising for future research. The comparison over time shows that terms like "deep learning" and "machine learning" have maintained their prominence, while more specialized terms like "variational autoencoders" and "unsupervised machine learning" are beginning to gain traction in recent discussions.

Figure 5, shown in Rojas Valdivia et al. [49], reveals a distinctive focus in the literature, with the prevalence of terms like "android", "detection", and "malware". This orientation highlights the practical application of Machine Learning algorithms in

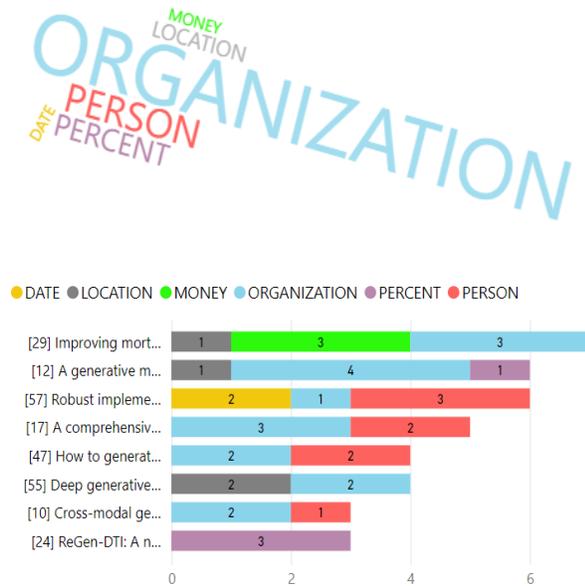


Fig. 7. NERs in abstracts by paper

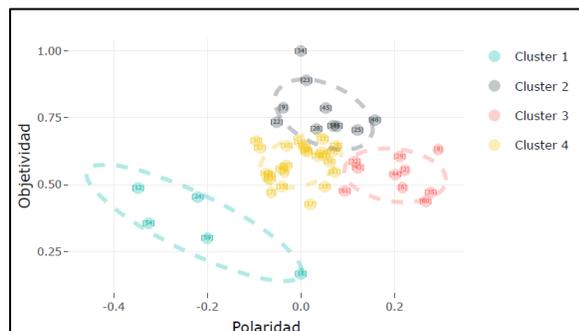


Fig. 8. Clusters of conclusions

mobile device security, demonstrating a commitment to solving real technological challenges.

The prevalence of terms such as "machine learning" and "generative models" suggests that these approaches are at the core of current research and may be key areas for future innovations. The diversification into more specific terms, such as "variational autoencoders" and "quantum machine learning," indicates an expansion towards more specialized and advanced approaches. This could imply that future research will focus on developing new applications and methodologies within generative learning and

its integration with emerging technologies like quantum computing.

**RQ3:** What sets of keywords tend to appear together in studies examining the Impact of Generative Algorithms on Machine Learning?

Figure 6 presents a bibliometric network focused on keywords, which shows the prevalent thematic connections in research on generative algorithms in Machine Learning.

The bibliometric network of keywords was generated using computational tools based on NLP and generative AI, allowing for the visualization of the relationship and co-occurrence of key terms in publications.

The clustering algorithm used, such as the Louvain algorithm for community detection, identifies groupings of related terms. In this case, keywords like "machine learning," "generative models," and "deep learning" form the most connected nodes, suggesting that these areas are highly interrelated in the reviewed research. The selection of keywords was based on key terms from the fields of machine learning and generative models, reflecting the central focus of research around these technologies. The strong connection between "machine learning" and "generative models" indicates that these techniques are frequently studied together.

Figure 11, from the study by Aparcana-Tasayco and Gamboa-Cruzado [5], places "machine learning" in a strategic position, linking it to specific challenges like network technology security. Alternatively, Figure 8 from Rojas Valdivia et al. [49], associates "machine learning" with advances in "artificial intelligence" and "deep learning", pointing towards applications in "internet of things" and "robotics".

The strong interrelationship between the key terms suggests that the field is converging towards greater integration of machine learning and generative techniques, which could accelerate advancements in areas such as artificial intelligence and complex data processing.

The presence of emerging approaches like "quantum machine learning" indicates a growing interest in exploring disruptive technologies that could transform the capabilities of generative models. However, the weaker connections suggest that these areas are still developing and

will require further research to consolidate within the main field.

**RQ4:** Which countries show co-occurrence in research on best practices and their impact on decision-making?

Figure 7 displays the frequencies of various named entities (NER) in the abstracts of papers, providing a detailed view of the main focuses of interest in this research field.

The figure analyzes how 'ORGANIZATION' dominates the abstracts, highlighting the fundamental role of institutions in the evolution of research on generative algorithms and Machine Learning. Additionally, the inclusion of 'PERSON' and 'PERCENT' illustrates the synergy between organizational support and individual contribution, along with the importance of statistical analysis.

Figure 11 in Rojas Valdivia et al. [49], shows a preference for 'PERCENT' in papers indexed in Scopus and ACM Digital Library, indicating an inclination towards studies with solid quantitative foundations. Contrasting with the current review, which highlights organizational support, the authors Rojas Valdivia et al. [49] emphasize the priority given to accuracy and measurability.

The frequent mention of the 'ORGANIZATION' category in the abstracts of the studies underlines the importance of institutional collaboration and support in the progress of research on generative algorithms and Machine Learning. Simultaneously, the prominent presence of 'PERSON' and 'PERCENT' indicates the relevance of individual contributions and the centrality of quantification and statistical analysis in research. These elements evidence that both individual expertise and analytical precision are fundamental in the study of generative algorithms and their impact on Machine Learning.

**RQ5:** What is the distribution of clusters of the conclusions of the papers distinguished by their high objectivity and minimal emotional polarity in the context of the Impact of Generative Algorithms on Machine Learning?

The cluster distribution illustrated in Figure 8 shows significant differences in the objectivity and polarity of the conclusions in the literature on generative algorithms in Machine Learning. This representation indicates variations in neutrality and

the emotional tone of the academic discourse, evidencing diversity in the approaches and perspectives adopted in the field.

The analysis of the clusters reveals variations in the approach and tone of the papers on generative algorithms and Machine Learning.

Cluster 1 stands out for its emotional neutrality, pointing to a rigorous technical analysis.

Cluster 2, with its trend towards positive conclusions, reflects an analytical approach that emphasizes favorable outcomes. In contrast,

Cluster 3 shows cautious optimism, combining objectivity with a slight positive tendency.

Cluster 4, on the other hand, displays diversity in objectivity and polarity, indicating a possible fusion of analysis and persuasion.

This spectrum of approaches not only enriches the understanding of the field but also underscores the plurality of perspectives in the study of these topics.

This study is distinguished by its novelty, given the scarcity of direct comparatives in the realm of generative algorithms and Machine Learning. It highlights the variability in methods and approaches used in research, contributing to a broader and deeper understanding of the area. The high objectivity observed in some clusters underscores their methodological robustness and the validity of their conclusions. The trend towards positive conclusions in certain groups reflects an inherent optimism about the potential of these algorithms, suggesting promising paths for future inquiries and underlining the importance of this work in advancing knowledge.

The variability between clusters reflects different styles in the interpretation of results, ranging from more neutral and objective to more subjective and optimistic. This may suggest the existence of more conservative approaches in certain research, while others may lean towards bolder claims. For future studies, understanding this distribution of conclusions can help identify trends in the presentation of results and how these may influence the perceived impact of the study. Additionally, it may be useful to consider how the degree of polarity and objectivity affects acceptance or critical appraisal by the scientific community.

## 5 Conclusions and Future Research

In this Systematic Literature Review, the impact of generative algorithms on Machine Learning (ML) is analyzed. A total of 62 papers were reviewed for the research, highlighting the growing relevance and dynamics of this field of study. The research emphasizes the importance of international collaboration, with the United States and Canada as leaders, alongside the active participation of Italy, China, and the United Kingdom.

The interconnection of key terms like "machine learning", "deep learning", and "generative model" in the bibliometric keyword network (RQ3) underscores the convergence of various methodologies and techniques, evidencing the interdisciplinary nature of the field. Moreover, the prominence of 'ORGANIZATION' in the paper abstracts (RQ4) emphasizes the crucial role of institutions in the research and development of the generative algorithms and ML field.

This trend highlights the importance of institutional support and the relevance of individual contributions and quantitative data, suggesting a balance between institutional cooperation and individual initiative in knowledge generation. In future research, we will incorporate quantitative metrics such as impact factor, citation count, and statistical rigor of the reviewed studies. This will complement the qualitative criteria, providing a more balanced evaluation and reducing potential interpretative biases.

## References

1. **Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J., Biswas, R. (2018).** Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Science and Technology*, 3(3), 1–16. DOI: 10.1088/2058-9565/aab859
2. **Fanfarillo, A. (2021).** Probabilistic forecasting using deep generative models. *Geoinformatica*, 25(1), 1–27. DOI: 10.1007/s10707-020-00425-8
3. **Sanchez-Lengeling, B., Aspuru-Guzik, A. (2018).** Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400), 360–365. DOI: 10.1126/science.aat2663
4. **Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R., Jensen, K. F. (2022).** Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5), e1608. DOI: 10.1002/wcms.1608
5. **Fan, C., Chen, M., Tang, R., Wang, J. (2022).** A novel deep generative modeling-based data augmentation strategy for improving short-term building energy predictions. *Building Simulation*, 15, 1–15. DOI: 10.1007/s12273-021-0807-6
6. **Yao, C., Cai, D., Bu, J., Chen, G. (2017).** Pre-training the deep generative models with adaptive hyperparameter optimization. *Neurocomputing*, 266, 499–507. DOI: 10.1016/j.neucom.2017.06.062
7. **Dahl, C. M., Sørensen, E. N. (2022).** Neural networks generative models for time series. *Journal of King Saud University - Computer and Information Sciences*, 34(2), 155–167. DOI: 10.1016/j.jksuci.2022.06.010
8. **Carvajal-Patiño, D., Ramos-Pollán, R. (2022).** Synthetic data generation with deep generative models to enhance predictive tasks in trading strategies. *Research in International Business and Finance*, 60, 101578. DOI: 10.1016/j.ribaf.2022.101578
9. **Paz, D., Zhang, H., Xiang, H., Liang, A., Christensen, H. I. (2023).** Conditional Generative Models for Dynamic Trajectory Generation and Urban Driving. *Sensors*, 23(15), 6764. DOI: 10.3390/s23156764
10. **Neo, E. R. K., Low, J. S. C., Goodship, V., Coles, S. R., Debattista, K. (2023).** Cross-modal generative models for multi-modal plastic sorting. *Journal of Cleaner Production*, 413, 137568. DOI: 10.1016/j.jclepro.2023.137568
11. **Alhomayani, F., Mahoor, M. H. (2021).** On oversampling imbalanced data with deep conditional generative models. *Expert Systems with Applications*, 165, 113856. DOI: 10.1016/j.eswa.2020.113856
12. **Shao, F., Shao, H., Wang, D., Lam, W. H. K., Cao, S. (2023).** A generative model for

- vehicular travel time distribution prediction considering spatial and temporal correlations. *Physica A: Statistical Mechanics and its Applications*, 616, 128646. DOI: 10.1016/j.physa.2023.128646
13. Carrillo-Perez, F., Pizurica, M., Ozawa, M. G., Vogel, H., West, R. B., Kong, C. S., Herrera, L. J., Shen, J., Gevaert, O. (2023). Synthetic whole-slide image tile generation with gene expression profile-infused deep generative models. *Cell Reports Methods*, 3(7), 100471. DOI: 10.1016/j.crmeth.2023.100471
  14. Fernandez Alcazar, F., Perdomo-Ortiz, A., Ghazi Vakili, M., Kalayci, C. (2022). Enhancing Generative Models via Quantum Correlations. *Physical Review X*, 12(2), 021037. DOI: 10.1103/PhysRevX.12.021037
  15. Felsch, G., Ghavidelnia, N., Schwarz, D., Slesarenko, V. (2023). Controlling auxeticity in curved-beam metamaterials via a deep generative model. *Computer Methods in Applied Mechanics and Engineering*, 403, 115646. DOI: 10.1016/j.cma.2023.115646
  16. Rongier, G., Rude, C., Herring, T., Pankratius, V. (2019). Generative Modeling of InSAR Interferograms. *Earth and Space Science*, 6(3), 533. DOI: 10.1029/2018EA000533
  17. Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review*, 37, 100317. DOI: 10.1016/j.cosrev.2020.100317
  18. Buzhinsky, I., Nerinovsky, A., Tripakis, S. (2023). Metrics and methods for robustness evaluation of neural networks with generative models. *Machine Learning*, 112(5), 1456–1478. DOI: 10.1007/s10994-021-05994-9
  19. Imrie, F., Bradley, A. R., van der Schaar, M., Deane, C. M. (2020). Deep Generative Models for 3D Linker Design. *Journal of Chemical Information and Modeling*, 60(4), 1983–1995. DOI: 10.1021/acs.jcim.9b01120
  20. Alcazar, J., Leyton-Ortega, V., Perdomo-Ortiz, A. (2020). Classical versus quantum models in machine learning: insights from a finance application. *Machine Learning: Science and Technology*, 1(3), 035003. DOI: 10.1088/2632-2153/ab9009
  21. Yoon, J., Mizrahi, M., Ghalaty, N., Jarvinen, T., Ravi, A., Brune, P., Kong, F., Anderson, D., Lee, G., Meir, A., Bandukwala, F., Kanal, E., Arik, S., Pfister, T. (2023). EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records. *npj Digital Medicine*, 6(1), 1–12. DOI: 10.1038/s41746-023-00888-7
  22. Kim, J.-Y., Cho, S.-B. (2022). Obfuscated malware detection using deep generative model based on global/local features. *Computers & Security*, 112, 102510. DOI: 10.1016/j.cose.2021.102510
  23. Ye, J. C. (2023). Unsupervised learning of global factors in deep generative models. *Pattern Recognition*, 137, 109179. DOI: 10.1016/j.patcog.2022.109179
  24. Sivangi, K. B., Amilpur, S., Dasari, C. M. (2023). ReGen-DTI: A novel generative drug target interaction model for predicting potential drug candidates against SARS-COV2. *Computational Biology and Chemistry*, 104, 107872. DOI: 10.1016/j.compbiolchem.2023.107872
  25. Sase, K., Shibuta, Y. (2023). Prediction of microstructure evolution at the atomic scale by deep generative model in combination with recurrent neural networks. *Acta Materialia*, 252, 118910. DOI: 10.1016/j.actamat.2023.118910
  26. Yang, L. M., Grooms, I. (2021). Machine learning techniques to construct patched analog ensembles for data assimilation. *Journal of Computational Physics*, 439, 110377. DOI: 10.1016/j.jcp.2021.110377
  27. Liu, L., Wang, S., Hu, B., Qiong, Q., Wen, J., Rosenblum, D. S. (2018). Learning structures of interval-based Bayesian networks in probabilistic generative model for human complex activity recognition. *Pattern Recognition*, 81, 545–561. DOI: 10.1016/j.patcog.2018.04.004
  28. Regenwetter, L., Srivastava, A., Gutfreund, D., Ahmed, F. (2023). Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design.

- Computer-Aided Design, 158, 103401. DOI: 10.1016/j.cad.2023.103401
29. **Hameed, M. A. B., Alamgir, Z. (2022).** Improving mortality prediction in acute pancreatitis by machine learning and data augmentation. *Computers in Biology and Medicine*, 148, 105911. DOI: 10.1016/j.combiomed.2022.105911
  30. **Marschall, M., Wübbeler, G., Schmähling, F., Elster, C. (2023).** Generative models and Bayesian inversion using Laplace approximation. *Computational Statistics*, 38, 2023–2048. DOI: 10.1007/s00180-023-01345-5
  31. **Hoh, M., Schöttl, A., Schaub, H., Wenninger, F. (2022).** A generative model for anomaly detection in time series data. *Procedia Computer Science*, 201, 1–8. DOI: 10.1016/j.procs.2022.03.001
  32. **Ramstead, M. J., Seth, A., Hesp, C., Sandved-Smith, L., Mago, J., Lifshitz, M., Pagnoni, G., Smith, R., Dumas, G., Lutz, A., Friston, K., Constant, A. (2022).** From generative models to generative passages: A computational approach to (neuro)phenomenology. *Review of Philosophy and Psychology*, 13, 25–47. DOI: 10.1007/s13164-021-00604-y
  33. **de Souza, M. D., Silva Junior, C. R., Quintino, J., Santos, A. L., da Silva, F. Q. B., Zanchettin, C. (2023).** Exploring the impact of synthetic data on human activity recognition tasks. *Procedia Computer Science*, 217, 1–10. DOI: 10.1016/j.procs.2023.01.001
  34. **Wang, M., Yu, J., Ning, Z.-H., Xiao, C.-B. (2021).** Compressed sensing using generative models based on Fisher information. *International Journal of Machine Learning and Cybernetics*, 12, 1–14. DOI: 10.1007/s13042-021-01337-1
  35. **Lu, M., Gibson, T. (2020).** Predictive and generative machine learning models for photonic crystals. *Nanophotonics*, 9(10), 1–12. DOI: 10.1515/nanoph-2020-0197
  36. **Yin, M., Zou, Z., Zhang, E., Cavinato, C., Humphrey, J. D., Karniadakis, G. E. (2023).** A generative modeling framework for inferring families of biomechanical constitutive laws in data-sparse regimes. *Journal of the Mechanics and Physics of Solids*, 172, 105200. DOI: 10.1016/j.jmps.2023.105200
  37. **Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V., Scotti, F. (2022).** A survey of unsupervised generative models for exploratory data analysis and representation learning. *ACM Computing Surveys*, 55(4), 1–32. DOI: 10.1145/3450963
  38. **Perraudin, N., Srivastava, A., Lucchi, A., Kacprzak, T., Hofmann, T., Réfrégier, A. (2019).** Cosmological N-body simulations: A challenge for scalable generative models. *Computational Astrophysics and Cosmology*, 6(1), 1–14. DOI: 10.1186/s40668-019-0032-1
  39. **Zaballa, O., Pérez, A., Gómez Inhiesto, E., Acaiturri Ayesta, T., Lozano, J. A. (2023).** Learning the progression patterns of treatments using a probabilistic generative model. *Journal of Biomedical Informatics*, 143, 104406. DOI: 10.1016/j.jbi.2023.104406
  40. **Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., Borrajo, L. (2023).** A survey on deep learning applied to medical images: From simple artificial neural networks to generative models. *Neural Computing and Applications*, 35(12), 893–917. DOI: 10.1007/s00521-022-07953-4
  41. **Zhao, P.-P., Zhu, H.-F., Liu, Y., Zhou, Z.-T., Li, Z.-X., Xu, J.-J., Zhao, L., Sheng, V. S. (2018).** A generative model approach for geo-social group recommendation. *Journal of Computer Science and Technology*, 33(5), 1058–1073. DOI: 10.1007/s11390-018-1852-1
  42. **Pagani, R. N., Kovalski, J. L., Resende, L. M. (2015).** Methodi Ordinatio: A proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. *Scientometrics*, 103(2), 485–507. DOI: 10.1007/s11192-015-1744-x
  43. **Wang, P., Xiao, J., Sun, C., Wu, X., Li, L., Yu, K. (2023).** Prediction of complex strain fields in concrete using a deep learning approach. *Construction and Building Materials*, 391, 131771. DOI: 10.1016/j.conbuildmat.2023.131771

44. **Mancisidor, R. A., Kampffmeyer, M., Aas, K., Jenssen, R. (2020).** Deep generative models for reject inference in credit scoring. *Knowledge-Based Systems*, 195, 105620. DOI: 10.1016/j.knosys.2020.105620
45. **Mancisidor, R. A., Kampffmeyer, M., Aas, K., Jenssen, R. (2022).** Generating customer's credit behavior with deep generative models. *Knowledge-Based Systems*, 239, 107253. DOI: 10.1016/j.knosys.2022.107253
46. **Thomas, S. M., Lefevre, J. G., Baxter, G., Hamilton, N. A. (2021).** Characterization of tissue types in basal cell carcinoma images via generative modeling and concept vectors. *Computerized Medical Imaging and Graphics*, 90, 101929. DOI: 10.1016/j.compmedimag.2021.101929
47. **Banian, S. (2019).** Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics*, 38(4), 1–15. DOI: 10.1145/3306346.3322971
48. **Borysov, S. S., Rich, J., Pereira, F. C. (2019).** How to generate micro-agents? A deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies*, 105, 66–83. DOI: 10.1016/j.trc.2019.05.020
49. **Lenz, S., Hess, M., Binder, H. (2021).** Deep generative models in DataSHIELD. *BMC Medical Research Methodology*, 21(1), 1–14. DOI: 10.1186/s12874-021-01237-6
50. **Otten, S., Caron, S., de Swart, W., van Beekveld, M., Hendriks, L., van Leeuwen, C., Podareanu, D., Ruiz de Austri, R., Verheyen, R. (2021).** Event generation and statistical sampling for physics with deep generative models and a density information buffer. *Nature Communications*, 12, 7985. DOI: 10.1038/s41467-021-22616-z
51. **Taniguchi, T., Nakamura, T., Suzuki, M., Kuniyasu, R., Hayashi, K., Taniguchi, A., Horii, T., Nagai, T. (2020).** Neuro-SERKET: Development of integrative cognitive system through the composition of deep probabilistic generative models. *New Generation Computing*, 38(4), 213–229. DOI: 10.1007/s00354-019-00084-w
52. **Iraki, T., Link, N. (2022).** Generative models for capturing and exploiting the influence of process conditions on process curves. *Journal of Intelligent Manufacturing*, 34, 225–240. DOI: 10.1007/s10845-021-01846-4
53. **Trappenberg, T. P. (2018).** A quantum machine learning algorithm based on generative models. *Science Advances*, 4(7), eaat9004. DOI: 10.1126/sciadv.aat9004
54. **Parr, T., Friston, K., Pezzulo, G. (2023).** Generative models for sequential dynamics in active inference. *Cognitive Neurodynamics*, 17(1), 87–105. DOI: 10.1007/s11571-023-09963-x
55. **Xue, T., Wallin, T. J., Menguc, Y., Adriaenssens, S., Chiamonte, M. (2020).** Machine learning generative models for automatic design of multi-material 3D printed composite solids. *Extreme Mechanics Letters*, 35, 100624. DOI: 10.1016/j.eml.2020.100624
56. **Krauth, T., Lafage, A., Morio, J., Olive, X., Waltert, M. (2023).** Deep generative modelling of aircraft trajectories in terminal maneuvering areas. *Machine Learning with Applications*, 13, 100281. DOI: 10.1016/j.mlwa.2022.100281
57. **Hwang, U., Kim, S.-W., Jung, D., Kim, S., Lee, H., Seo, S. W., Seong, J.-K., Yoon, S. (2023).** Real-world prediction of preclinical Alzheimer's disease with a deep generative model. *Artificial Intelligence in Medicine*, 136, 102168. DOI: 10.1016/j.artmed.2023.102168
58. **Leyton-Ortega, V., Perdomo-Ortiz, A., Perdomo, O. (2021).** Robust implementation of generative modeling with parametrized quantum circuits. *Quantum Machine Intelligence*, 3(1), 9. DOI: 10.1007/s42484-021-00040-2
59. **Winci, W., Buffoni, L., Sadeghi, H., Khoshaman, A., Andriyash, E., Amin, M. H. (2020).** A path towards quantum advantage in training deep generative models with quantum annealers. *Machine Learning: Science and Technology*, 1(4), 045023. DOI: 10.1088/2632-2153/aba220
60. **Liu, W., Zhang, Y., Deng, Z., Zhao, J., Tong, L. (2019).** A conditional generative model based on quantum circuit and classical optimization. *International Journal of*

- Theoretical Physics, 58, 1923–1936. DOI: 10.1007/s10773-019-04005-x
61. **Feng, X., Zhao, Q., Ma, J., Jiang, G. (2020).** On modeling and predicting popularity dynamics via integrating generative model and rich features. *Expert Systems with Applications*, 158, 113538. DOI: 10.1016/j.eswa.2020.113538
  62. **Li, Y., Yao, Y., Xia, Y., Tang, M. (2023).** Searching for protein variants with desired properties using deep generative models. *BMC Bioinformatics*, 24(1), 100. DOI: 10.1186/s12859-023-05415-9
  63. **Han, Z.-Y., Wang, J., Fan, H., Wang, L., Zhang, P. (2018).** Unsupervised generative modeling using matrix product states. *Physical Review X*, 8(3), 031012. DOI: 10.1103/PhysRevX.8.031012
  64. **Kitchenham, B., Charters, S. (2007).** Guidelines for performing systematic literature reviews in SE. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, 1–64. DOI: 10.1145/12345678
  65. **Zhang, E. Y., Cheok, A. D., Pan, Z., Cai, J., Yan, Y. (2023).** From Turing to Transformers: A comprehensive review and tutorial on the evolution and applications of generative transformer models. *Scientific Reports*, 13(1), 7895. DOI: 10.1038/s41598-023-33416-6
  66. **Zhang, A., Xing, L., Zou, J., Wu, J. C. (2022).** Shifting machine learning for healthcare from development to deployment and from models to data. *Nature Biomedical Engineering*, 6(10), 1213–1227. DOI: 10.1038/s41551-022-00898-y
  67. **Anstine, D. M., Isayev, O. (2023).** Generative models as an emerging paradigm in the chemical sciences. *Journal of the American Chemical Society*, 145(4), 1890–1902. DOI: 10.1021/jacs.2c13467
  68. **Cárdenas-Quispe, A., Vergaray-Mezarina, R., Gamboa-Cruzado, J. (2021).** Machine learning para la detección de malware en Android: Revisión sistemática de la literatura. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 2021(E28), 505–519.
  69. **Aparcana-Tasayco, A. J., Gamboa-Cruzado, J. (2022).** Machine learning for management in software-defined networks: A systematic literature review. *IEIE Transactions on Smart Processing and Computing*, 11(6), 400–412. DOI: 10.5573/IEIESPC.2022.11.6.400
  70. **Rojas Valdivia, J., Gamboa-Cruzado, J., de la Cruz Vélez de Villa, P. (2023).** Systematic literature review on machine learning and its impact on APIs deployment. *Computación y Sistemas*, 27(4), 1301–1315. DOI: 10.13053/CyS-27-4-4371.

*Article received on 12/01/2025; accepted on 21/04/2025.  
\*Corresponding author is Flavio Amayo-Gamboa.*