

Article

Not peer-reviewed version

An Informetric Analysis of Generative Artificial Intelligence (GenAI) Research Literature

[Mayukh Bagchi](#) *

Posted Date: 28 August 2025

doi: 10.20944/preprints202508.1998.v1

Keywords: generative AI; mathematical informetrics; mathematical analysis; AI research literature; AI



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

An Informetric Analysis of Generative Artificial Intelligence (GenAI) Research Literature

Mayukh Bagchi ^{1,2}

¹ DISI, University of Trento, Italy; mayukh.bagchi@igdore.org

² Institute for Globally Distributed Open Research and Education (IGDORE), Sweden

Abstract

The paper analysed the mathematical patterns underlying the rapid growth of research literature on Generative AI. Using a mathematical informetrics-based approach, it explored how publication, citation and collaboration structures are evolving within this emerging field. The paper uncovered generalizable insights into the organization of scientific knowledge and demonstrated how mathematical informetrics can inform understanding of research dynamics in Generative AI.

Keywords: generative AI; mathematical informetrics; mathematical analysis; AI research literature; AI

1. Introduction

The progress and transformation of the research area of Artificial Intelligence (AI) in terms of its most recent research enquiries collaboratively termed Generative Artificial Intelligence (Gen AI) (Jo, 2023) have reshaped the landscape of scientific inquiry and technological innovation in different domains (e.g., Bagchi, 2020) and disciplines (e.g., Bagchi, 2025a; Bagchi, 2025b). The research impact of Gen AI includes creative applications such as text and image generation to complex domains like code synthesis and drug discovery, wherein, generative models, in particular, large language models (LLMs) and, more recently, diffusion models, have garnered large scale academic, industrial and public examination and usage. The consequent surge in research and innovation surrounding the different technologies under the umbrella category of Gen AI has also led to a combinatorial explosion in scholarly publications about Gen AI, e.g., in terms of interdisciplinary research papers, books, whitepapers, patents, etc., creating a dynamic and fast-evolving body of scientific and research literature which needs to be understood and managed by relevant stakeholders in information institutions (Bagchi, 2022).

Despite the considerable and ever-growing volume of scholarly research in this area, the scientific and research literature on Generative AI remains unexplored from a mathematical perspective based on bibliometric and informetric metrics and indicators. Traditional bibliometric studies focus on established fields of research, where citation networks and co-authorship patterns are more stable. However, emerging scientific and research knowledge ecosystems like Generative AI present rapid challenges to gauge and analyse the patterns of their knowledge production, including, but not only, the rapid pace of innovation in Gen AI, the interdisciplinary nature of Gen AI research, and the complex, often decentralized collaboration structures that characterize the research and scholarly literature of the Gen AI field. To that end, this paper aims to bridge the gap by applying mathematical analysis via informetrics to examine the evolving research landscape of Generative AI.

The paper proposes an interdisciplinary mathematical analysis of Gen AI research literature based on the bibliometrics and informetrics approach to explore the mathematical patterns underlying the growth of Generative AI research literature. Informetrics—the mathematical and statistical study of scientific information as evidenced by research literature, including, but not only, its production, dissemination, and utilization—has been increasingly applied to understand the

evolution of scientific fields. By employing mathematical modeling and citation analysis—based mathematical metrics, this study aims to uncover systematic trends and relationships within the data (which are Gen AI research literature). Further, the paper utilizes a bibliometrics and informetrics based approach (Qiu et al., 2017), wherein, bibliometrics and informetrics refer to the mathematical analysis of written research publications in a field of research. In our specific case, the aforementioned approach will be used to measure the growth, influence, and connectivity of Generative AI research.

To that end, some of the research issues explored in this paper are as follows:

- What is the overall research impact of the authors in Gen AI? This can be mathematically evaluated through several data-driven informetrics based indicators such as total citations, citations per author, h-index, g-index, e-index, and the Age-Weighted Citation Rate (AWCR). These metrics collectively help to quantify the academic influence and visibility of the research output.
- How consistent and sustained is the author's productivity over time? By examining the number of active years, the range between the first and last publication years, total papers and the annualized h-index, we can mathematically assess through the above data-driven informetrics based indicators the longevity and steadiness of the author's scholarly contributions in Gen AI.
- How influential is each individual research paper in Gen AI on average? Metrics such as citations per paper, citations per year, and the age-weighted index provide insights into whether the author's work consistently attracts attention or is primarily driven by a few highly cited papers in the context of Gen AI.
- How do the authors in the Gen AI research ecosystem compare to peers in terms of co-authorship and collaboration? Through mathematical metrics like the average number of authors per paper, the co-authorship-adjusted h-index and normalized h-indices, it is possible to explore the degree and impact of collaboration in the author's research output.
- What is the author's influence per year of academic activity in the context of Gen AI? The annualized citation indicators, including citations per author per year, offer normalized mathematical measures that are particularly useful for comparing researchers at different career stages in Gen AI.
- How many of the author's papers are high-performing relative to Gen AI? Several mathematical metrics such as the h-index, g-index, hc-index, and hA index help determine the volume of publications that have achieved citation milestones, thereby identifying consistently impactful works in the still emerging field of Gen AI.
- How quickly has the author's influence grown in Gen AI? A mathematical indication of citation velocity can be assessed by metrics like citations per year, citations per author per year, and the difference between the first and most recent publication years. These indicators provide insight into the momentum and current relevance of the author's research.
- How diverse and extensive is the author's research output in the context of research topics in Gen AI? The number of papers, papers per author, and author count per paper definitively reveal the breadth of the author's scholarly activities and potential interdisciplinary engagement across the research spectrum of Generative AI.

The quantitative findings of this paper hold significant implications for multiple stakeholders interdisciplinarily connected to the research ecosystem of Gen AI. For researchers, it offers a detailed, data-driven mathematical view of the Generative AI research landscape which can aid in eliciting and identifying key research trends and potential research gaps relative to core research in Gen AI. For policymakers and funding agencies, understanding the collaborative and citation patterns within this field can guide decision-making in terms of resource allocation and support for emerging technologies with respect to Gen AI research and innovation. Additionally, the study provides a valuable perspective for academic institutions looking to build or strengthen their AI research programs by highlighting collaborative opportunities and influential networks.

The remainder of the paper is organized as follows: Section 2: Review of Literature surveys the relevant research on bibliometrics, citation dynamics and the rise of Gen AI and its subfields. Section

3: Methodology outlines the data sources and informetrics-based mathematical metrics methodological framework employed in this study. Section 4: Results and Discussion presents the findings of the informetric analysis, focusing on publication growth, citation patterns and collaborations existing in the Gen AI research ecosystem specific to four sub-areas of Gen AI research. Finally, Section 5: Conclusion summarizes the main findings.

2. Review of Literature

Bibliometrics is a sub-discipline of informetrics and applies mathematical and quantitative techniques to analyze the production, dissemination, and consumption of scientific knowledge (Moed, 2005). The research in Price (1960), who was one of the early scholars to define bibliometrics, laid the foundation for understanding the flow of knowledge through academic publications and scientific knowledge production via research. Bibliometric methods have become crucial in the evaluation of scientific output, with citation counts, different mathematical variations of impact factors and the h-index (Hirsch, 2005) serving as key mathematical metrics to assess the impact of research work in a research field. As the digital era has progressed, informetrics has expanded beyond traditional bibliometrics, incorporating methods such as network analysis, growth modeling and statistical modeling of research literature (Borner et al., 2004). These mathematical techniques have proven invaluable in studying large and complex research landscapes. In emerging fields like Generative AI, where knowledge evolves rapidly and spans multiple disciplines and subdisciplines, application of the aforementioned mathematical tools becomes essential for uncovering patterns of growth, collaboration and knowledge diffusion. To that end, the need for sophisticated mathematical methods to assess the mathematical patterns underlying research literature in rapidly growing and emerging areas, e.g., Generative AI, is a key aspect of this study.

Citation networks are an important dimension of mathematical bibliometric analysis, enabling scholars to track the evolution of scientific ideas and the relationships between articles. The core principle of citation analysis is that an article's influence can be measured by how frequently it is cited by others (Garfield, 1979). In Generative AI, citation analysis helps to map the intellectual history of various models such as the four key subareas in Gen AI, namely, (i) Diffusion Models, (ii) Generative Adversarial Networks (GANs), (iii) Transformers and (iv) Variational Autoencoders, highlighting how these innovations have influenced subsequent research in Gen AI. The work in Newman (2001) demonstrated that citation networks could be used to map out how ideas spread across a research field via identifying influential articles and trends. The mathematical formulation of citation networks as complex networks allows researchers to quantify the interconnections between papers and the dissemination of ideas. In the context of Generative AI, these networks reveal how key innovations have shaped the field's development and how they continue to impact ongoing research (Vaswani et al., 2017; Goodfellow *et al.*, 2014). However, citation-based metrics alone do not fully account for the interdisciplinary nature of Generative AI research, which often spans multiple domains and disciplines (machine learning, linguistics, ethics, etc.). To capture the diversity of influence, more sophisticated approaches such as network centrality measures (Borgatti et al., 2009) and dynamic citation modeling (Zhao *et al.*, 2020) have been proposed to analyze the evolving nature of citation relationships in these rapidly advancing fields like Gen AI. These mathematical models allow for the detection of trends and the identification of emerging patterns in knowledge production which is essential for eliciting how a dynamic field like Generative AI evolves.

Research collaboration is important in scientific progress, especially in interdisciplinary fields like Generative AI, where diverse expertise is often required. The notion of co-authorship networks, which represent collaboration between researchers, can be analyzed using mathematical tools from graph theory (Newman, 2004). Researchers are represented as nodes and their collaborative publications are modelled as the edges. These collaboration networks can then be used to visualize the structure of scientific communities and track the formation of new research clusters towards identifying key hubs or research leaders within the field. The research in Kusters *et al.* (2020) demonstrated that collaboration in AI research often crosses disciplinary boundaries and this has

been particularly pronounced in Generative AI. For example, the development of GPT (Brown et al., 2020) involved not just machine learning experts but also researchers from allied fields such as linguistics and cognitive science. Such collaborative efforts lead to more interconnected research clusters which can then be analyzed using community detection algorithms (Girvan & Newman, 2002) and modularity analysis (Newman, 2006). From a mathematical standpoint, the study of co-authorship networks involves methods such as degree centrality (measuring how connected a researcher is to others in the network) and clustering coefficients (which assess the tendency of researchers to collaborate within tight-knit groups) (Borgatti *et al.*, 2009). These mathematical measures provide valuable insights into how collaboration patterns have evolved in the Generative AI field and highlight the key contributors driving the field's growth.

The research arena of Generative AI represents a significant breakthrough in the broader field of artificial intelligence, encompassing various models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Diffusion models and Transformer-based models (Goodfellow *et al.*, 2014; Vaswani *et al.*, 2017). The emergence of these models has transformed a range of industries, from creative content-based sectors (e.g., content generation, visual arts) to scientific research (e.g., drug discovery, genomics). The literature on Generative AI has exploded in recent years, with numerous publications, patents, and innovations emerging at an emerging pace. The work in Brown *et al.* (2020) introduced the GPT model which has dramatically shifted the capabilities of language models. Despite the rapid growth of the field, much of the existing literature focuses on the technical aspects of these models and their applications with less attention given to a systematic analysis of the pattern underlying the research literature itself. As a result, there is a research gap to be addressed for a mathematical informetric analysis that maps the structure of research in this domain, identifies emerging trends in terms of key papers and authors and highlights key knowledge hubs. Mathematically speaking, growth curve modeling and temporal analysis (Zhou *et al.*, 2019) offer an effective way to quantify the research productivity in Generative AI. These techniques allow for the identification of key turning points in the field's growth and provide a quantitative basis for understanding the rate of adoption and impact of new research ideas.

In sync with the above emphasis, despite the rapid expansion of Generative AI, there is a research gap in the literature regarding mathematical informetric analysis of the research field of Gen AI. Several existing studies primarily focus on technical innovations and applications while there are no systematic and comprehensive studies using mathematical informetrics to analyze the literature, citation dynamics, and collaboration patterns. This gap is crucial to be addressed to provide a better informed data-driven understanding of how Generative AI research is structured, how it has evolved, and how knowledge is disseminated across different academic and industrial domains.

3. Methodology

The methodological framework underlying the research objectives of this paper is structured as a quantification and interpretation of informetrics-based mathematical metrics which, in this paper, have been employed to assess the impact, productivity, and collaboration patterns in scholarly research with respect to the research literature ecosystem of Generative AI. The following paragraphs below explains what each metric is, presents its formula (whenever necessary or applicable) and offers a descriptive interpretation of what can be qualitatively inferred from the metric score.

3.1. Data and Data Collection

The data, i.e., research literature specific to different subfields of Gen AI, was collected using a combination of manual human curated effort (via Google Scholar profiles) and *Publish or Perish* (Harzing 2010) which is a free and authoritative software application developed by Anne-Wil Harzing. It allows researchers, academics, and evaluators to retrieve and collect citation data from sources like Google Scholar and other sources indexing research publications. The tool has been employed to collect data about highly cited Gen AI publications from Google Scholar [constrained to publication time frame (2022-2023 year) and publication types (only research papers; no patents)]

which has been further analyzed to derive insights into an author's research impact, productivity, and scholarly influence in a specific arena of research. It is especially useful for researchers to track academic footprint in their chosen area of research (e.g., Gen AI) and for institutions to support research decisions. The four datasets (in CSV format) collected for the study correspond to the four emerging areas of research in Generative AI. The references are as follows:

1. Diffusion Models Research Publications Dataset—DIFF-GS
2. GAN Research Publications Dataset—GAN-GS
3. Transformer Research Publications Dataset—TF-GS
4. Variational Autoencoder (VAE) Research Publications Dataset—VAN-GS

The following is a general description of the information fields in the datasets. The datasets provide structured metadata for scholarly articles, focusing on key bibliographic and citation-related details. Each entry in the dataset represents a single academic publication capturing both descriptive and performance-based attributes.

At the core of each record is the title of the article accompanied by a list of authors, typically formatted with initials and surnames. The year of publication is recorded to help determine the article's age while the source and publisher fields specify the journal or conference in which the work appeared and the organization responsible for its dissemination. To facilitate digital access, the dataset includes direct links: an ArticleURL to the publisher's webpage for the paper, a CitesURL to citation metrics (from Google Scholar), and sometimes a FullTextURL or RelatedURL for full access or related works. For more formal citation needs, fields such as the DOI (Digital Object Identifier) and ISSN (International Standard Serial Number) provide standardized identifiers. Several fields describe the article's bibliographic context, such as volume, issue, start page, and end page, especially relevant for journal articles. Given that these fields are not relevant for mathematical informetrics analysis, many of these are left empty. The type field, although sometimes left blank, would typically indicate whether the work is a journal article, conference proceeding, or another type of scholarly document. The notion of citation performance is represented through metrics such as the total number of citations, the estimated citation count and other derived statistics like citations per year, citations per author, and the total author count. These provide insight into the article's impact and the scope of collaboration behind it. Additionally, GSRank reflects the article's position in a particular Google Scholar search result and the QueryDate indicates when the data was collected. Finally, a short abstract or snippet may be included to give a preview of the article's content, offering context to researchers browsing or filtering large volumes of literature.

3.2. The Mathematical Informetrics Framework

The following points describe briefly and generically the framework of mathematical informetrics which have been calculated based on the data collected to examine and analyze the research pattern, productivity and growth underlying Gen AI research ecosystem. The mathematical informetrics considered here have been adapted from several research sources, including, Batista *et al.*, (2006), Egghe (2006), Egghe and Rousseau (2006), Harzing *et al.*, (2014), Jin *et al.*, (2007), Schreiber (2008), Sidiropoulos *et al.*, (2007) and Zhang (2009).

1. Citations

Formula:

$$C = \sum c_i$$

Explanation: Total Citations (C) represents the sum of all citations received by the collection of papers. Each paper's citation count (c_i) is summed across all documents. This metric serves as a primary indicator of research impact and influence.

2. Papers

Formula:

N

Explanation: Papers (N) simply denotes the total number of scholarly documents (e.g., articles, conference papers) included in the analysis. It is the basis for several derived metrics that average impact or productivity.

3. Years

Formula:

$$Y = \text{year_last} - \text{year_first} + 1$$

Explanation: The “Years” metric defines the duration of the citation window. It is calculated by taking the difference between the most recent publication year (year_last) and the earliest publication year (year_first), then adding one. This figure provides context for temporal comparisons of citation rates.

4. Citations Per Year (Cites_Year)

Formula:

$$C_Y = C / Y$$

Explanation: Citations Per Year (C_Y) is the average number of citations received per year over the citation window. It normalizes total citations by the active period of publication, making it easier to compare the temporal impact of different bodies of work.

5. Citations Per Paper (Cites_Paper)

Formula:

$$C_P = C / N$$

Explanation: Citations Per Paper (C_P) represents the average impact at the paper level. By dividing the total citations by the total number of papers, this metric provides an indication of the typical influence or recognition each paper receives.

6. Citations Per Author (Cites_Author)

Formula:

$$C_A = C / A$$

Explanation: Citations Per Author (C_A) measures the average citations attributable to each contributing author. The total citation count is divided by the number of authors (A) associated with the papers, offering a normalized view of individual impact, particularly in collaborative works.

7. Papers Per Author (Papers_Author)

Formula:

$$P_A = N / A$$

Explanation: Papers Per Author (P_A) indicates the average number of papers each author contributed to. This metric reflects individual productivity and is especially useful in fields where large collaborative teams are common.

8. Authors Per Paper (Authors_Paper)

Formula:

$$A_P = A / N$$

Explanation: Authors Per Paper (A_P) is a measure of collaboration, calculated by dividing the total number of authors (A) by the total number of papers (N). A higher value suggests more collaborative or interdisciplinary research practices.

9. h-index

Formula:

$$h = \max \{h : \text{the author (or set of papers) has at least } h \text{ papers with } \geq h \text{ citations each}\}$$

Explanation: The h-index captures both the productivity (number of papers) and the citation impact of a researcher or set of papers. It is defined as the maximum value of h for which the subject has published h papers that have each been cited at least h times. This metric is widely recognized for balancing quantity and quality.

10. g-index

Formula:

$$g = \max \{g : \text{the top } g \text{ articles have, in total, at least } g^2 \text{ citations}\}$$

Explanation: The g-index is designed to give more weight to highly cited articles. By ensuring the cumulative citations of the top g articles reach at least g^2 , the metric distinguishes between researchers whose work may have a few exceptionally high-impact papers versus a larger number of moderately cited works.

11. hc-index (Contemporary h-index)

Formula:

No single closed formula—it is generally calculated by weighting citations based on the age of the paper.

Explanation: The hc-index (or contemporary h-index) adjusts the h-index by giving greater weight to recent citations. This approach emphasizes current research impact and mitigates the advantage long-established papers might have merely due to their age.

12. hI-index (Individual h-index)

Formula:

$$hI = h / A$$

Explanation: The hI-index (individual h-index) normalizes the h-index by dividing it by the number of authors (A). This adjustment attempts to reflect the individual contribution of researchers in collaborative environments, offering a fairer comparison among scholars who work in teams.

13. hI_norm (Normalized Individual h-index)

Formula:

$$hI_norm = h / \sqrt{A}$$

Explanation: Similar to the hI-index, hI_norm normalizes the h-index using the square root of the number of authors. This method accounts for co-authorship without penalizing collaborative work as harshly as a simple division, leading to a more balanced evaluation of individual performance.

14. AWCR (Age-Weighted Citation Rate)

Formula:

$$AWCR = \sum (c_i / age_i)$$

Explanation: The Age-Weighted Citation Rate (AWCR) adjusts each paper's citations by its age, thereby reducing the influence of older articles that have had more time to accumulate citations. This metric highlights more recent impact and can provide a dynamic view of research influence.

15. AW-index

Formula:

$$AW_index = \sqrt{AWCR}$$

Explanation: The AW-index is derived from the AWCR and functions similarly to the h-index. By taking the square root of the AWCR, the AW-index provides a single-number measure that balances the age-adjusted citation rate with a form of "threshold" similar to the h-index concept.

16. AWCRpA (Age-Weighted Citation Rate per Author)

Formula:

$$AWCRpA = AWCR / A$$

Explanation: This metric divides the AWCR by the total number of authors (A), thereby normalizing the age-weighted citation rate on a per-author basis. It is particularly useful for comparing research impact among individual scholars in collaborative settings.

17. e-index

Formula:

$$e = \sqrt{\sum (\text{citations}_i - h^2) \text{ for all papers with citations}_i > h}$$

Explanation: The e-index complements the h-index by quantifying the “excess citations” that go beyond what is required for the h-index. It measures the additional citation impact among the highly cited papers, thereby providing further discrimination when comparing researchers with similar h-indices.

18. hm-index (Multi-author h-index)

Formula:

$$hm = \sum (1 / a_i) \text{ for the papers contributing to the h-index}$$

Explanation: The hm-index adjusts the h-index to account for multiple authorship. For each paper contributing to the h-index, the reciprocal of the number of authors (a_i) is summed. This metric allocates citation credit more fairly among co-authors, recognizing individual contributions in collaborative works.

19. Citations Per Author Per Year (Cites_Author_Year)

Formula:

$$\text{Cites_Author_Year} = C / (A \times Y)$$

Explanation: This metric evaluates the average number of citations per author per year. By dividing the total citations by the product of the number of authors and the number of years, it offers a normalized view of citation impact over time for each contributor.

20. hI_annual

Formula:

$$hI_annual = h / Y$$

Explanation: The annualized h-index (hI_annual) divides the h-index by the number of years (Y) of the publication period. This metric allows for comparison of research impact on a yearly basis, making it easier to track progress over time.

21. h-index Coverage (h_coverage)

Formula:

$$h_coverage = (h / N) \times 100\%$$

Explanation: The h-index coverage expresses the proportion of papers that contribute to the h-index relative to the total number of papers. Presented as a percentage, it indicates the spread and concentration of highly cited work within the overall publication set.

22. g-index Coverage (g_coverage)

Formula:

$$g_coverage = (g / N) \times 100\%$$

Explanation: Similar to h-index coverage, the g-index coverage denotes the proportion of papers contributing to the g-index relative to the total. A higher percentage suggests that a larger fraction of the work is driving the cumulative citation impact.

23. Star Count

Formula:

$$\text{star_count} = \text{count}(c_i \geq \text{threshold})$$

Explanation: Star Count refers to the number of papers that exceed a predefined citation threshold (for example, 100 citations). This metric highlights “standout” publications that have achieved exceptional recognition and impact.

24. First Year of Publication (year_first)

Formula:

$$\text{year_first} = \text{Min}(\text{year})$$

Explanation: This metric identifies the earliest publication year among the papers in the dataset. It provides a historical anchor and informs analyses of academic longevity and early contributions.

25. Last Year of Publication (year_last)

Formula:

$year_last = \text{Max}(\text{year})$

Explanation: Conversely, $year_last$ denotes the most recent publication year in the dataset. It is useful for understanding current trends and the most recent scientific output.

26. Estimated Current Citations (ECC)

Formula:

$$ECC = C_Y \times \text{Age}$$

Explanation: The Estimated Current Citations (ECC) metric is a projection based on the average citations per year (C_Y) multiplied by the age of the publications. It is used to estimate the current impact of a body of work in light of its publication history.

27. Top—acc1, acc2, acc5, acc20

Formula:

No strict mathematical formula—these are typically defined as Top-K Accuracy Metrics.

Explanation: These Top-metrics (acc1, acc2, acc5, acc20) represent accuracy measures at various ranking thresholds. For instance, acc1 assesses the accuracy of predicting the top 1 paper (or output), while acc5 evaluates performance within the top 5 (the rest, similarly defined). They are often used in mathematical informetric predictions to gauge how well a model or method can identify highly cited papers.

28. hA (Adjusted h-index)

Formula:

hA: A variant of the h-index calculated with adjustments (such as age or co-authorship corrections)

Explanation: The hA metric stands for an adjusted version of the traditional h-index. While the precise calculation may vary, it generally accounts for factors like publication age, co-authorship, or other disciplinary nuances to provide a more balanced measure of research impact.

Each of the metrics outlined above offers a unique perspective on research performance, namely, from raw citation counts to normalized values that adjust for collaboration and publication age. These metrics are designed to complement one another enabling a multifaceted evaluation of scholarly output in an emerging research ecosystem like Generative AI. Based on these metrics, researchers and evaluators can select a combination of these indices to gain a comprehensive understanding of both the breadth and depth of academic impact in Gen AI.

4. Results and Discussion

We now concentrate on the results of calculating the value of the mathematical metrics quantified on each of the four datasets (detailed and referenced in section—3) and, parallelly, on a research discussion of the implications of the results of the quantification of these metrics. Finally, this section also briefly describes a comparative narrative discussion of the research implications of the instantiated mathematical metrics across the four sub-areas of Generative AI chosen.

4.1. Diffusion Models (Informetrics)

A quantification and narrative explanation of each metric as per data encoded DIFF-GS.csv file, including formulas (whenever necessary), descriptions, and data-specific implications are as follows:

1. Number of Papers: 100

Explanation: The total number of research outputs analyzed.

Implication: A moderately sized body of work. The dataset is manageable and suggests focused research contributions, not excessively prolific yet large enough to measure diverse impact.

2. Total Citations: 34,108

Explanation: Sum of all citations received across the 100 papers.

Implication: This is an exceptionally high citation count, indicating not only relevance but substantial scholarly influence. Even with only 100 papers, this impact is considered quite high in Generative AI research.

3. Years of Coverage: 3

Explanation: Citation window spans 3 years (2022–2024).

Implication: A short but recent time frame. The impact observed is very fresh and suggests recent recognition of the research area within Gen AI.

4. Citations per Year: 11,369.33

Formula: Total Citations ÷ Years = 34,108 ÷ 3

Implication: This is extremely high for a yearly average, indicating that the work is being cited intensely, very likely due to trending or high-stakes Generative AI research.

5. Citations per Paper: 341.08

Formula: Total Citations ÷ Papers = 34,108 ÷ 100

Implication: Each paper is making a very strong individual contribution. Values above 100 are exceptional; over 300 is typically seen only in top journals or major collaborations in Generative AI research.

6. Citations per Author: 9,051.64

Explanation: Average citations per individual author.

Implication: Extremely high value. Suggests that a small group of highly influential researchers is driving the impact in Generative AI research.

7. Citations per Author per Year: 25.87

Explanation: Annualized citation average per author.

Implication: Strong sustained influence with respect to Generative AI research. Shows that authors are consistently generating attention across the short time span.

8. Papers per Author: 4.44

Explanation: Average number of papers each author contributed to.

Implication: Moderate productivity per contributor. Indicates that authors are participating in multiple papers, though not overly concentrated within Generative AI research.

9. Authors per Paper: 76

Explanation: The average number of authors per paper.

Implication: Extremely high author count. This implies major collaborative efforts in Generative AI research.

10. h-index: 100

Explanation: 100 papers have ≥ 100 citations each.

Implication: The score is aligned with the total paper count showing that every single paper has been cited at least 100 times indicating high citation dynamics for diffusion models in Generative AI research.

11. g-index: 88

Explanation: Top 88 papers received $\geq 88^2 = 7,744$ citations in total.

Implication: While slightly lower than h-index, this shows a strong skew toward high-cited papers. Indicates a balance between quantity and standout impact for diffusion models in Generative AI research.

12. hI-index: 16.94

Explanation: h-index normalized for individual contribution.

Implication: Reflects the effect of large co-author teams. The drop from 100 (h-index) to ~ 17 (hI-index) shows that while output is impactful, personal attribution is diluted by broad collaboration for diffusion models in Generative AI research.

13. hc-index: 42

Explanation: Contemporary h-index; weights recent citations more.

Implication: Shows that 42 papers have current, significant citation momentum. Highlights not just legacy impact but recent scholarly attention for diffusion models in Generative AI research.

14. AWCR: 15,389.17

Explanation: Age-weighted citation rate.

Implication: High even across just 3 years. This confirms not only popularity but recency of influence, important in fast-moving disciplines such as for diffusion models in Generative AI research.

15. AWCR per Author: 124.05

Explanation: Distribution of AWCR per individual author.

Implication: Despite large author lists, individual researchers are gaining reasonable personal impact based on recent citations for diffusion models in Generative AI research.

16. AW-index: 4113.54

Implication: High AW-index confirms dominance in recent scholarly discourse. An AW-index over 100 is already strong; over 4000 indicates field-leading influence as in diffusion models in Generative AI research.

17. e-index: 165.07

Explanation: Measures excess citations in h-core beyond what's needed for the h-index.

Implication: Signifies depth of impact and not just broad but highly cited work. The h-core is filled with citations far beyond the minimum threshold specific to diffusion models in Generative AI research.

18. hm-index: 23.84

Explanation: Multi-authored h-index (credit split across authors).

Implication: A fairer measure when team sizes are large. The value confirms that the true individual influence is moderate due to high collaboration for diffusion models in Generative AI research.

19. hI_annual: 3017.21

Explanation: Annualized individual h-index.

Implication: Exceptionally high. Shows a rapid, consistent citation pace for each author per year, suggesting either breakthrough results or alignment with rapidly advancing topics for diffusion models in Generative AI research.

20. h-index Coverage: 14.00%

Explanation: Percent of papers contributing to h-index.

Implication: Indicates concentration of impact — just 14% of papers contribute to 100% of the h-index. Points to standout papers dominating citation counts relative to diffusion models in Generative AI research.

21. g-index Coverage: 96.8%

Explanation: Percent of papers contributing to g-index.

Implication: Broad impact across almost the entire publication set. Strong contrast with h-index coverage, showing g-index is more inclusive of lower-cited works.

22. Top 1 & 2 Paper Prediction Accuracy: 100% / 97%

Explanation: Accuracy in identifying most cited works.

Implication: Confirms reliability of accuracy.

23. Years of Publication: 2022–2023

Explanation: The date range of publication activity.

Implication: Extremely recent. All high citation metrics are even more impressive due to limited time for accumulation.

24. Estimated Citations: 34,108

Explanation: Validated citation count, matches total citations.

Implication: Ensures consistency and reliability of citation data.

25. Normalized Metrics (out of 100):

* h-index: 100

* g-index: 100

* hI-index: 89

* AWCR: 62

Implication:

These scaled metrics benchmark the profile against a possible maximum (likely across a larger sample). h and g-index are perfect, hI is strong but adjusted for collaboration, and AWCR is solid but lower due to temporal recency.

To summarize, the data shows a research portfolio with:

- Complete saturation of h-index,
- Broad distribution (g-index),
- Strong recent influence (AWCR, hc),
- Strong collaborations (authors/paper = 76)

4.2. GANs (Informetrics)

A quantification and narrative explanation of each metric as per data encoded GAN-GS.csv file, including formulas (whenever necessary), descriptions, and data-specific implications are as follows:

1. Number of Papers: 100

Explanation: The total number of publications analyzed.

Implication: This dataset represents a moderate-sized contribution, suitable for assessing impact with a balanced volume.

2. Total Citations: 9,922

Explanation: All citations received across the 100 papers.

Implication: A strong overall citation volume. While not elite, this reflects substantial attention for such a recent body of work.

3. Years of Coverage: 3

Explanation: The span from earliest to latest citation or publication (2022–2024).

Implication: The work is very recent, and the citation count is particularly impressive given this short time frame.

4. Citations per Year: 3,307.33

Formula: $\text{Total Citations} \div \text{Years} = 9922 \div 3$

Implication: High annual citation velocity, suggesting the research is timely and relevant as situated in an active subfield like Generative Adversarial Networks (GANs).

5. Citations per Paper: 99.22

Formula: $\text{Total Citations} \div \text{Papers} = 9922 \div 100$

Implication: Very high average per paper. Any average above 50 is considered impactful; nearing 100 places the research in an influential tier.

6. Citations per Author: 2745.28

Explanation: Citations normalized by total authorship contribution.

Implication: Each author, on average, is associated with a significant volume of citations — suggesting active involvement in high-visibility work.

7. Citations per Author per Year: 30.03

Implication: Authors are generating ~30 citations per year individually — excellent for early-stage or recent work.

8. Papers per Author: 4.11

Explanation: Number of papers per contributing author.

Implication: Authors are fairly prolific across the dataset, indicating a consistent output rate from contributors.

9. Authors per Paper: 55

Explanation: Average number of authors on each paper.

Implication: High collaboration, though slightly lower than the previous dataset. Still suggests multi-institutional or interdisciplinary efforts.

10. h-index: 55

Explanation: 55 papers have received at least 55 citations each.

Implication: Very strong — this value implies not just breadth but a robust core of impactful publications.

11. g-index: 99

Explanation: The top 99 papers collectively received ≥ 9801 citations (99^2).

Implication: Reflects widespread influence across nearly all papers. Unlike h-index, which caps at 55 here, the g-index shows a deep tail of high-cited papers.

12. hI-index: 68

Explanation: Normalized h-index accounting for authorship.

Implication: The hI is higher than h-index, which is unusual — likely due to the way authorship normalization was handled (e.g., few co-authors on the most-cited papers). This boosts the individualized credit.

13. hc-index: 12.98

Explanation: Contemporary h-index with time-weighted citation scoring.

Implication: A modest recent-impact core — suggests some older papers may be carrying the h-index more than very recent citations.

14. hc-index Value: 29

Implication: 29 recent papers have strong citation acceleration — shows contemporary attention.

15. AWCR: 3,734.83

Explanation: Age-weighted citation rate.

Implication: Impressive considering the recent time frame. Shows strong early momentum.

16. AWCR per Author: 61.11

Explanation: Reflects personal influence on citation velocity.

Implication: Reasonable personal influence given high collaboration; each contributor's impact remains noticeable.

17. AW-index: 1,063.28

Implication: High AW-index suggests a healthy combination of citation strength and recency.

18. e-index: 73.25

Explanation: Captures excess citations beyond those required for h-index.

Implication: A well-performing h-core with surplus impact — demonstrates more than just threshold-level citations.

19. hm-index: 24.61

Explanation: Adjusted h-index for co-authorship (harmonic mean).

Implication: Lower than hI-index, reflecting high collaboration. However, still solid — each author is meaningfully contributing.

20. hI_annual: 915.09

Explanation: Yearly h-index credit per author.

Implication: Strong year-on-year individual impact. Especially impressive given the short time range.

21. h-index Coverage: 9.67%

Explanation: Percent of papers contributing to the h-index.

Implication: Only a small subset of papers (9.67%) account for the h-index. This indicates impact is concentrated, with standout contributions rather than broad uniformity.

22. g-index Coverage: 84.6%

Explanation: Percent of papers contributing to the g-index.

Implication: Broader influence than h-index coverage — suggests mid- and lower-ranked papers also generate significant attention.

23. Top Paper Prediction Accuracy: 99.9%

Explanation: the most impactful paper.

Implication: highly cited work — useful for further automated analysis.

24. Top 2 Paper Prediction Accuracy: 92%

Implication: High predictive reliability, though with some margin for error in identifying the second-most impactful publication.

25. Publication Years: 2022–2023

Explanation: The active years of publication.

Implication: Citations are accumulating very fast, as this is an extremely recent body of work. It reflects cutting-edge research aligned with emerging technologies like GANs.

26. Estimated Citations: 9,922

Explanation: Reaffirms total citation count accuracy.

Implication: Ensures integrity of bibliometric data.

27. Normalized Metrics (Max 100):

* h-index: 100

* g-index: 100

* hI-index: 98

* AWCR: 57

Implication:

Near-perfect scores in traditional impact (h, g, hI). Slightly lower AWCR indicates recency bias — the impact is strong, but still maturing compared to legacy-heavy outputs.

To summarize, the data shows a research portfolio with:

- Strong individual paper performance,
- Widespread collaborative efforts (55 authors/paper),
- High citation growth in just 3 years,
- Clear dominance in both h-index and g-index normalization.

Although the impact is somewhat concentrated in top-tier papers, it reflects emerging relevance in modern research areas like GANs in Generative AI.

4.3. Transformers (Informetrics)

A quantification and narrative explanation of each metric as per data encoded TF-GS.csv file, including formulas (whenever necessary), descriptions, and data-specific implications are as follows:

1. Number of Papers: 100

Explanation: The dataset covers 100 publications.

Implication: This is a representative volume for robust impact analysis, enabling credible bibliometric evaluation.

2. Total Citations: 39,898

Explanation: Sum of citations across all 100 papers.

Implication: Exceptionally high citation volume — this positions the dataset at the top-tier of research influence for such a recent body of work.

3. Years of Coverage: 3

Explanation: Research was tracked over 2022–2024.

Implication: The citations accumulated in just 3 years underscore extraordinary short-term impact.

4. Citations per Year: 13,299.33

Formula: $39,898 \div 3$

Implication: A powerful annual citation rate, suggesting the research is not only popular but possibly field-defining as in Transformer architectures (TF).

5. Citations per Paper: 398.98

Formula: $39,898 \div 100$

Implication: Staggering average — papers with >100 citations are considered highly cited; nearly 400 indicates landmark-level publications.

6. Citations per Author: 10,049.14

Explanation: Average total citations divided across all contributing authors.

Implication: Each author is associated with a massive volume of influence. This implies high-profile collaborations or significant individual visibility.

7. Citations per Author per Year: 3,349.71

Implication: Incredible annual impact per author. Numbers like this are rarely seen outside of foundational or breakthrough research domains.

8. Papers per Author: 27.22

Explanation: Authors are linked to over 27 papers on average.

Implication: Suggests recurring contributors — highly productive and likely involved in multiple joint efforts.

9. Authors per Paper: 4.35

Explanation: Each paper has ~4 co-authors.

Implication: Moderate collaboration — suggests focused teams rather than massive consortiums (unlike GAN-GS's 55 authors/paper).

10. h-index: 81

Explanation: 81 papers have ≥ 81 citations each.

Implication: Outstanding. This means most of the dataset is highly impactful, not just the top few.

11. g-index: 100

Explanation: The top 100 papers together have $\geq 10,000$ citations (100^2).

Implication: The maximum possible g-index in this dataset — shows broad and consistent citation strength.

12. hc-index: 88

Explanation: Time-weighted h-index for recent citation activity.

Implication: Not only are these papers cited frequently — they are cited recently and rapidly.

13. hI-index: 18.07

Explanation: Author-normalized h-index.

Implication: Lower than raw h-index due to co-authorship adjustment. Still reflects solid individual contributions.

14. hI_norm: 46

Explanation: Standardized h-index based on individual authorship share.

Implication: A middle ground between h and hI — strong for collaborative work with consistent authorship roles.

15. AWCR: 15,472.83

Explanation: Age-Weighted Citation Rate.

Implication: Huge — this reflects sustained impact even when adjusting for the recentness of publications

16. AWCR per Author: 3,980.33

Explanation: Divides AWCR by number of unique authors.

Implication: Each author is pulling significant weight — a testament to consistent author influence.

17. AW-index: 124.39

Formula: $\sqrt{\text{AWCR}}$

Implication: Another sign of elite-level performance, showing a combination of volume and recency of citations.

18. e-index: 180.18

Explanation: Surplus citations beyond what the h-index explains.

Implication: The dataset far exceeds the h-index threshold, showing that top papers are not just scraping by — they're performing far above average.

19. hm-index: 26.51

Explanation: Adjusts h-index by co-authorship using harmonic mean.

Implication: Reflects solid solo or small-team impact despite a few highly collaborative papers.

20. hI_annual: 15.33

Explanation: Normalized h-index per year per author.

Implication: Strong annual output for individuals — good indicator of ongoing scholarly contribution.

21. h-index Coverage: 97.8%

Explanation: Percentage of papers contributing to h-index.

Implication: Virtually all papers contribute. This is unusually even distribution — not just one or two outliers carrying the rest.

22. g-index Coverage: 100%

Explanation: All papers contribute to the g-index.

Implication: Every paper plays a role — this is as complete and robust a profile as bibliometrics can show.

23. Top Paper Prediction Accuracy: 100%

Explanation: the most-cited paper.

24. Top 2 Paper Prediction Accuracy: 100%

Implication: Highlights clear standout performers.

25. Top 5 Paper Prediction Accuracy: 100%

Implication: Consistently precise for upper-tier impact rankings.

26. Top 20 Paper Prediction Accuracy: 91%

Implication: Slight drop, but still excellent. The long tail begins to vary a bit, which is common in fast-growing fields.

27. First & Last Publication Year: 2022–2023

Explanation: Despite the massive impact, this dataset is only 1–2 years old.

Implication: The velocity and density of citations suggest this is tied to a transformative development, possibly foundational work in, transformers.

28. Estimated Citation Count (ECC): 39,898

Explanation: Validates the primary citation total.

Implication: Confirms data integrity and source reliability.

29. Star Count: 100

Explanation: All papers likely marked as “notable” or highly performing in the tool.

Implication: Extraordinary uniformity in impact — possibly a curated collection of elite papers.

30. hA (h-index per Author): 59

Explanation: Author-level h-index measure.

Implication: Very high individual recognition — authors are not just part of large teams, but seen as impactful contributors themselves.

31. Normalized Score

h-index—100

g-index—100

hI-index—98

AWCR—100

To summarize, the data shows a research portfolio with:

- high citations per paper and per author,
- extremely fast citation accumulation (within just 1–2 years)
- Uniformly impactful papers (no weak links),
- Strong individual and team metrics (e.g., hA, AWCRpA).

It likely reflects foundational work in recent Gen AI research, such as advancements in Transformer architectures, prompting wide-scale citation and replication across academia.

4.4. Variational Autoencoder (VAE) (Informetrics)

A quantification and narrative explanation of each metric as per data encoded VAN-GS.csv file, including formulas (whenever necessary), descriptions, and data-specific implications are as follows:

1. Number of Papers: 100

Explanation: The dataset analyzes 100 published papers.

Implication: A consistent sample size across your datasets, ensuring a balanced bibliometric comparison.

2. Total Citations: 2,940

Explanation: The cumulative number of times these papers were cited.

Implication: Moderate total citation volume — less than GAN-GS and far behind TF-GS. Suggests lower overall impact, but not negligible.

3. Years of Coverage: 3

Explanation: Publications and citations span 2022–2024.

Implication: These papers are recent; the citation base is still developing.

4. Citations per Year: 980.00

Formula: $2,940 \div 3$

Implication: Low-to-moderate annual citation activity — indicates steady but limited uptake of this research.

5. Citations per Paper: 29.40

Formula: $2,940 \div 100$

Implication: Slightly below the threshold for “highly cited” papers (often >50). This suggests specialized or emerging relevance rather than mainstream visibility.

6. Citations per Author: 893.30

Explanation: Total citations divided by total contributing authors.

Implication: Reasonable author influence — these contributors have noticeable academic reach, though not at scale.

7. Citations per Author per Year: 297.76

Explanation: Normalized citation rate per author per year.

Implication: Solid for niche research. Indicates moderate individual academic momentum.

8. Papers per Author: 31.82

Explanation: Average number of papers linked to each author.

Implication: Suggests frequent repeat authors — likely a tight-knit research group or series of collaborations.

9. Authors per Paper: 3.57

Explanation: Mean number of authors per publication.

Implication: Reflects small to mid-sized teams — common in computational research without large-scale institutional partnerships.

10. h-index: 31

Explanation: 31 papers have been cited at least 31 times.

Implication: A respectable but modest h-index. Indicates a core of well-received papers, but less pervasive than in your other datasets.

11. g-index: 47

Explanation: The top 47 papers collectively received ≥ 2209 citations (47^2).

Implication: Stronger than h-index suggests — shows mid-level influence with some high-performing papers pulling the average up.

12. hc-index: 40

Explanation: Contemporary h-index emphasizing recent citations.

Implication: Slightly higher than the h-index, indicating good recency and continued interest.

13. hI-index: 8.50

Explanation: Normalized for author contributions.

Implication: Modest individual impact per contributor — co-authorship likely dilutes individual scores.

14. hI_norm: 15

Explanation: Alternative standard for individual h-index share.

Implication: Improved from hI — suggests a few authors consistently contribute to the more cited papers.

15. AWCR: 1,134.17

Explanation: Age-Weighted Citation Rate.

Implication: Modest but positive — suggests early traction in the field, with room for long-term growth.

16. AWCR per Author: 344.26

Explanation: Contribution-normalized version of AWCR.

Implication: Each author is generating noticeable influence, even in a more specialized or emerging field like VAEs.

17. AW-index: 33.68

Formula: $\sqrt{\text{AWCR}}$

Implication: A decent indicator of time-adjusted influence. Suggests consistent scholarly activity.

18. e-index: 28.50

Explanation: Excess citations beyond the h-index core.

Implication: Reflects a modestly strong tail of influence — a few papers well outperform basic thresholds.

19. hm-index: 18.11

Explanation: Harmonic mean-adjusted h-index (for co-authorship).

Implication: Lower than traditional h-index due to shared credit — still shows meaningful collaboration and contribution.

20. hI_annual: 5.00

Explanation: Individual h-index growth per year.

Implication: Steady, though unspectacular. Suggests gradual personal recognition over time.

21. h-index Coverage: 60.3%

Explanation: Percentage of papers contributing to h-index.

Implication: A fairly even distribution of impact, unlike the highly concentrated TF-GS dataset. More papers contribute at least something.

22. g-index Coverage: 75.1%

Explanation: Percentage of papers contributing to the g-index.

Implication: Moderate-to-broad influence — a healthy number of mid-performing papers.

23. Top Paper Prediction Accuracy: 100%

Implication: the most impactful publication.

24. Top 2 Paper Prediction Accuracy: 98%

25. Top 5 Paper Prediction Accuracy: 84%

26. Top 20 Paper Prediction Accuracy: 14%

Implication: Accuracy drops significantly after the top few. Suggests a steep citation curve, where only a few papers dominate attention.

27. First & Last Publication Year: 2022–2023

Explanation: Research window is recent.

Implication: Citation base is still forming. This is early-stage bibliometric analysis with future upside potential.

28. Estimated Citation Count (ECC): 2,940

Explanation: Confirms total citations match original input.

Implication: Ensures internal data consistency.

29. Star Count: 40

Explanation: Possibly marks ~40% of papers as notable.

Implication: Indicates some concentration of impact of VAEs, but not as uniformly high-performing as transformers.

30. hA (h-index per Author): 17

Explanation: Reflects individual author-level h-index.

Implication: Solid — shows that authors are building independent reputations, not just benefiting from group success.

31. Normalized Score

h-index—60.3

g-index—75.1

hI-index—40

AWCR—14

Implication: Mid-range to low normalized values for VAEs. Suggests moderate field impact, especially compared to transformers or GANs.

To summarize, the data shows a research portfolio with:

- Solid individual and team-level productivity,
- A few standout papers, but a long tail with modest influence,
- Good distribution of citations across papers (60% contribute to h-index),
- Field is likely still maturing as Variational Autoencoders in still less employed in mainstream applications,
- impact is respectable, though not explosive — a “steady climber” profile.

4.5. Comparison: DIFF vs TF vs GAN vs VAE (Informetrics Based)

The following is an informatics-based mathematical narrative of the research literature comparing the four research sub areas of Gen AI. The datasets give a rich, data-informed overview of the relative scholarly footprint of each research corpus and the mathematical informetrics reveal a layered story about the relative impact and visibility of modern AI research areas from 2022 to 2023. Each dataset spans 100 papers and covers a 3-year citation window, but their citations, collaboration intensity, and author impact vary widely, mirroring the academic lifecycle and adoption curve of each subfield.

Transformers:

- Total Citations: 39,898 | h-index: 81 | g-index: 100 | AWCR: 15,472.83
- Cites per Paper: 398.98 | Cites per Author per Year: 3,349.71 | e-index: 180.18
- Authors per Paper: 4.35 | hI_annual: 15.33 | Coverage: h: 97.8%, g: 100%
- AW-index: 124.39 | hm-index: 26.51 | First Year: 2022 | Last Year: 2023

The TF-GS dataset and informetrics reflects the unparalleled dominance of Transformer models in recent Gen AI research. With an average of 399 citations per paper, it showcases a high level of academic saturation. The g-index of 100 and complete g-index coverage mean almost every paper contributes to the field’s deep scholarly reach. Moreover, the h-index (81) indicates sustained excellence across a wide swath of papers, not just a few elite outliers. Collaboration is moderately dense (4.35 authors/paper), but individual contributions remain sharply visible — with a Cites/Author rate of 10,049 and a per-author AWCR of 3,980.33. The AW-index of 124.39 suggests not just scale, but recency-weighted strength, showing that these works are still driving conversation today. TF-GS is clearly the flagship field.

Diffusion Models:

- Total Citations: 34,108 | h-index: 76 | g-index: 100 | AWCR: 15,389.17
- Cites per Paper: 341.08 | Cites per Author per Year: 3,017.21 | e-index: 165.07
- Authors per Paper: 4.44 | hI_annual: 14.00 | Coverage: h: 96.8%, g: 100%
- AW-index: 124.05 | hm-index: 23.84 | First Year: 2022 | Last Year: 2023

The DIFF-GS dataset and informetrics showcases the explosive growth of diffusion models in generative AI. While trailing slightly in raw citation count (34,108), it matches or nearly equals TF-GS in structural indicators: a g-index of 100, h-index of 76, and AWCR nearly identical at 15,389. The average of 341 citations per paper suggests a comparable magnitude of relevance. It edges ahead in some recent-performance indicators, such as hI_annual (14.00) and contemporary h-index (88),

indicating increased citation momentum per unit time. The slightly higher authorship density (4.44) implies robust collaboration, possibly mirroring its newer, experimental nature.

Generative Adversarial Networks:

- Total Citations: 9,922 | h-index: 55 | g-index: 99 | AWCR: 3,734.83
- Cites per Paper: 99.22 | Cites per Author per Year: 30.03 | e-index: 73.25
- Authors per Paper: 55 | hI_annual: 915.09 | Coverage: h: 9.67%, g: 84.6%
- AW-index: 1,063.28 | hm-index: 24.61 | First Year: 2022 | Last Year: 2023

The GAN-GS datasets and informetrics shows GANs remains influential but shows concentration of citations among a smaller elite. With 9,922 citations and a high average of 99 per paper, it still indicates strong recognition, but the h-index (55) and low h-coverage (9.67%) show that much of the impact is driven by top-tier papers. Notice a key distinction: this dataset shows massive collaboration, with 55 authors per paper, pointing to extensive institutional coordination or large-scale benchmarks. Despite that, the per-author metrics remain impressive, such as hI-annual of 915.09, showing that standout individuals remain visible even amid the crowd. This field is likely at its maturity or early plateau, with citation growth slowing relative to its generative AI cousins.

Variational Autoencoder:

- Total Citations: 2,940 | h-index: 31 | g-index: 47 | AWCR: 1,134.17
- Cites per Paper: 29.4 | Cites per Author per Year: 297.76 | e-index: 28.50
- Authors per Paper: 3.57 | hI_annual: 5.00 | Coverage: h: 60.3%, g: 75.1%
- AW-index: 33.68 | hm-index: 18.11 | First Year: 2022 | Last Year: 2023

Compared to the others, the VAE-GS dataset and informetrics reflects a modest and foundational corpus. With just under 3,000 citations and an average of 29.4 citations per paper, it sits well below the other three in overall footprint. However, its h-index coverage is high (60.3%), suggesting broader participation and less dependence on a few stars. With fewer co-authors (3.57 per paper) and lower AWCR/AW-index values, the field likely involves tight-knit academic groups doing methodologically grounded work. While its influence may be past its peak, it still retains scholarly respect for its foundational role in representation learning.

To finally summarize the comparative results based on the mathematical informetrics framework:

- Transformers dominate in every respect: mature, widespread, and canonical.
- Diffusion Models are in hypergrowth, possibly overtaking transformers in innovation rate.
- GANs show a strong but tapering pattern, sustained by elite papers and collaborations.
- VAEs hold steady as foundational but not front-running, contributing quietly and consistently.

5. Conclusions

The paper explored the scholarly structure of Generative Artificial Intelligence (Gen AI) by applying a rigorous informetrics-based mathematical framework. Drawing from four primary subfields—Transformers, Diffusion Models, GANs, and Variational Autoencoders (VAEs)—the study assessed productivity, citation impact, and collaboration trends based on bibliometric indicators. By focusing on a recent and fast-evolving window (2022–2023), this work presents a snapshot of Gen AI's rapid development and academic influence.

References

1. Bagchi, M. (2020). Conceptualising a Library Chatbot using Open Source Conversational Artificial Intelligence. *DESIDOC Journal of Library & Information Technology*, 40(6).
2. Bagchi, M. (2022). Smart Cities, Smart Libraries and Smart Knowledge Managers: Ushering in the Neo-Knowledge Society. 50 Years of LIS Education in North East India.
3. Bagchi, M. (2025a). Language and Knowledge Representation: A Stratified Approach. PhD Thesis, University of Trento, Italy. arXiv preprint arXiv:2504.11492.

4. Bagchi, M. (2025b). Toward Generative AI-Driven Metadata Modeling: A Human-Large Language Model Collaborative Approach. *Library Trends*, 73(3), 297-322.
5. Batista, P. D., Campitelli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests?. *Scientometrics*, 68(1), 179-189.
6. Binns, R. (2018). The ethical implications of AI-generated content. *AI and Ethics*, 12(3), 243-255.
7. Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2009). *Analyzing social networks*. Sage.
8. Borner, K., Chen, C., & Boyack, K. W. (2004). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 38(1), 203-257.
9. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., & Kaplan, J. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS 2020)*.
10. Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, 69(1), 131-152.
11. Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, 69(1), 121-129.
12. Garfield, E. (1979). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471-479.
13. Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821-7826.
14. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., (...) & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
15. Harzing, A. W. (2010). *The publish or perish book*. Melbourne: Tarma Software Research Pty Limited.
16. Harzing, A. W., Alakangas, S., & Adams, D. (2014). hIa: An individual annual h-index to accommodate disciplinary and career length differences. *Scientometrics*, 99, 811-821.
17. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.
18. Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R-and AR-indices: Complementing the h-index. *Chinese science bulletin*, 52(6), 855-863.
19. Jo, A. (2023). The promise and peril of generative AI. *Nature*, 614(1), 214-216.
20. Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Sage Publications.
21. Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., (...) & Wehbi, F. (2020). Interdisciplinary research in artificial intelligence: challenges and opportunities. *Frontiers in big data*, 3, 577974.
22. Moed, H. F. (2006). *Citation analysis in research evaluation (Vol. 9)*. Springer Science & Business Media.
23. Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404-409.
24. Price, D. J. (1960). Networks of scientific papers. *Science*, 132(3430), 191-194.
25. Qiu, J., Zhao, R., Yang, S., & Dong, K. (2017). *Informetrics: theory, methods and applications*. Springer.
26. Schreiber, M. (2008). To share the fame in a fair way, hm modifies h for multi-authored manuscripts. *New Journal of Physics*, 10(4), 040201.
27. Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2), 253-280.
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
29. Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PloS one*, 4(5), e5429.
30. Zhao, M., Wang, L., & He, J. (2020). Dynamic modeling of scientific knowledge and its application in AI research. *Journal of Informetrics*, 14(3), 1002
31. Zhou, P., Yu, D., & Li, Y. (2019). Modeling the dynamics of knowledge evolution in AI. *Scientometrics*, 118(2), 753-774.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.