A BIBLIOMETRIC ANALYSIS OF AI-DRIVEN PROCEDURAL CONTENT GENERATION IN VIDEO GAMES: KEY CONTRIBUTORS, THEMATIC TRENDS, AND COLLABORATION NETWORKS FROM 2008-2025

A BIBLIOMETRIC ANALYSIS OF AI-DRIVEN PROCEDURAL CONTENT GENERATION IN VIDEO GAMES: KEY CONTRIBUTORS, THEMATIC TRENDS, AND COLLABORATION NETWORKS FROM 2008-2025

Ralf William Alexander Schmidt

This Independent Study Manuscript was Presented to
The Graduate School of Bangkok University
in Partial Fulfillment
of the Requirements for the Degree
Master of Management in Business Innovation

Academic Year 2024

This manuscript has been approved by
the Graduate school
Bangkok University

Title: A Bibliometric Analysis of AI-Driven Procedural Content Generation in Video Games: Key Contributors, Thematic Trends, and Collaboration Networks from 2008-2025

Author: Ralf William Alexander Schmidt

Independent Study Committee:

Advisor: Dr. Ronald Vatananan-Thesenvitz

Field Specialist: Dr. Detlef Reis

Schmidt, R. W. A, Master of Management (Business Innovation), July 2025,
Graduate School, Bangkok University

A Bibliometric Analysis of AI-Driven Procedural Content Generation in Video Games: Key Contributors, Thematic Trends, and Collaboration Networks from 2008-2025 (42 pp.)

Advisor: Ronald Vatananan-Thesenvitz, Ph.D.

ABSTRACT

AI-driven Procedural Content Generation via Machine Learning (PCGML) is profoundly impacting the video game industry, yet its academic literature remains highly fragmented. This creates a significant challenge for researchers and practitioners needing to track trends, identify foundational work, and find expert collaborators. This study attempts to address this gap by providing a data-driven map of the PCGML research field through a rigorous bibliometric analysis.

Following PRISMA guidelines, this study analyzed 2,184 curated documents from the Scopus database (2008-2025). A quantitative analysis using the bibliometrix R package examined the field's conceptual, intellectual, and social structures. Revealing a clear evolution from classic AI toward a modern core dominated by deep learning and reinforcement learning. The United Kingdom was identified as the leader in research impact, while network analysis uncovered a core group of highly influential authors. This paper provides a valuable roadmap for academics by highlighting research gaps and offers industry practitioners strategic insights for technology adoption and collaboration, effectively bridging the gap between academic theory and practical application.

Keywords: Procedural Content Generation (PCG), Machine Learning, Artificial Intelligence (AI), Generative AI, PCGML, Deep Learning, Neural Networks, Reinforcement Learning, Computational Creativity, Video Games, Game Development, Game Design, Level Design, Collaboration Networks, Trend Analysis, Citation Analysis

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my advisor for this Independent Study, Dr. Ronald Vatananan-Thesenvitz. His guidance, insightful feedback, and profound patience have been essential throughout every stage of this independent study. His expertise in research methodology (especially Bibliometrix) was invaluable in shaping the direction and ensuring the quality of this analysis, especially during a research journey that required significant time and iteration.

My sincere thanks go to the Graduate School of Bangkok University for providing the institutional support and academic environment necessary to complete this work.

I also want to acknowledge my colleagues in the gaming industry, whose real-world challenges and rapid adoption of generative AI provided the initial spark and sustained motivation for this project.

Finally, I wish to thank my family and friends for their unwavering encouragement and moral support throughout this journey. Their belief in me was a constant source of strength.

Ralf William Alexander Schmidt

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

This chapter lays the groundwork for the independent study (IS) by establishing the context, objectives, and significance of conducting a bibliometric analysis of the AI-driven Procedural Content Generation (PCGML) research field. As a rapidly evolving and interdisciplinary domain, a structured overview is crucial for navigating the complex landscape of PCGML. This introduction will guide the reader from the broad context of AI's role in the modern video game industry to the specific research problem that this study addresses.

The current chapter begins by providing essential background and context, defining the key concepts: *Procedural Content Generation (PCG)* and its modern version, *machine-learning-driven evolution (PCGML),* furthermore, the Problem Statement will be addressed and explained together with the study's primary Research Objective and Questions. Finally, the Scope, Significance of the Study, and Definition of Terms are presented to clarify the boundaries of the analysis and highlight its expected contributions to academia, the game industry and society at large.

1.1 Background and Context of the Study

Since its origins as a niche hobby with games like Tennis for Two in 1958, the global video game industry has transformed into a dominant cultural and economic force through undergoing rapid growth, with a projected market value exceeding $300 billion by 2027 (Statista, 2023). A key driver of this expansion is the emergence and use of artificial intelligence (AI), which is broadly defined as a set of systems that can reason, learn, and act autonomously (Russell & Norvig, 2021). Within the industry, AI enables a non-player character (NPC) to independently navigate a complex environment, an enemy to adapt its combat tactics in response to player actions, or a system to dynamically adjust a game's difficulty (Millington & Funge, 2009). While these established uses are critical, this study focuses on one of AI's most transformative current applications, utilizing it not just to populate game worlds in the form of belivable non player characters (NPCs), but to create the environments and worlds themselves through AI-driven Procedural Content Generation (PCG) and Procedural Content Generation via Machine Learning (PCGML)

1.1.1 Introduction: PCG and PCGML

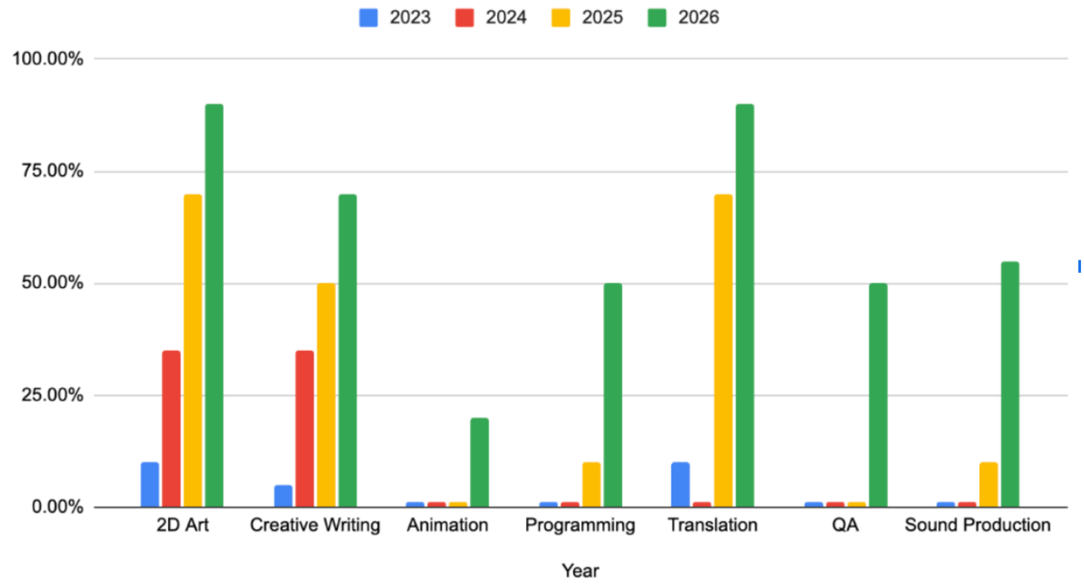
PCG, at its core, refers to the algorithmic creation of game content (such as levels, environments, in-game items, and even narratives) with a limited need for direct or indirect human input (Short & Adams, 2017). As opposed to requiring manual design and content generation, developers establish a set of rules and parameters used by an AI agent that utilize those rules and parameters to generate vast quantities of unique content. This new AI-driven way of procedurally generate content allows for increased replayability for players, reduced manual labor costs, and the ability to generate game worlds on a scale previously unfeasible using manual labor alone (Hendriks, 2023).

Historically, PCG has mostly relied on simpler, hand-coded algorithms and rule-based systems to generate content (Yannakakis & Togelius, 2018), however, recent advances in machine learning have made a shift toward Procedural Content Generation via Machine Learning (PCGML), by leveraging techniques such as Generative Adversarial Networks (GANs) and Reinforcement Learning (RL) (Volz et al., 2018; Jain, Isaksen, & Togelius, 2020). This modern approach has given rise to a more complex and rapidly growing body of literature, and to navigate this field of research and understand its key concepts, challenges, and breakthroughs, this study will employ a bibliometric analysis of the research field.

1.2 Problem Statement

The motivation for this study stems from direct professional experience within the video game industry since 2020. Over the last three years in the industry, I have led the adoption of generative AI into our company's core development pipeline, witnessing firsthand its transformation from a theoretical tool to an essential component in our daily workflows for 2D art, creative writing, ideation, and many more applications. Our internal data illustrates this dramatic adoption, with the proportion of labor handled by AI projected to grow from under 10% in 2023 to over 75% by 2025 (as seen by Figure 1.1 below).

Figure 1.1: Figure Depicting the Percentage Adoption of AI in the Author's Workplace Between 2023 and 2026 (Predicted).



While this internal data is anecdotal, it serves to illustrate the real-world problem that motivated this study. The rapid, hands-on adoption experienced in the company has highlighted a critical gap between industry practice and academic knowledge. The central problem is that the scholarly landscape of PCGML and other AI tools is highly fragmented and undocumented, which became a significant obstacle as we sought to refine our AI workflows and adopt new ones. Knowledge of which techniques are most effective and which academic labs are producing foundational work remains scattered across hundreds of sources, making it difficult for us, and potentially, other practitioners to navigate the current state of the field.

The problem reveals a disconnect between academic achievements and industry application. For developers and industry leaders, this fragmentation creates a barrier to innovation, making it difficult to make informed, evidence-based decisions on technology adoption and to identify expert collaborators, which then naturally also means that for the academic community, it increases the risk of redundant research and makes it harder for new researchers to identify meaningful gaps experienced in the practical application of their research. The consequence of not addressing this could naturally result in slower progress and maturation of the field, potentially hindering the practical and strategic implementation of transformative AI technologies in the industrial sector, and society at large.

This study is a direct response to this real-world need, aiming to bridge the gap between those that seek, and those that provide, new innovative discoveries of PCGML by generating the resource needed: a clear, evidence-based map of the PCGML research field.

1.3 Research Objective and Questions

In order to address the previously discussed gap between academia and the industry with regards to PCGML, the primary objective of this study is to systematically map the academic, intellectual and social structure of the AI-driven PCGML research field, from 2008 to 2025, by identifying the related core themes, key contributors, and collaborative dynamics through a rigorous Bibliometric study. In order to to so, the following sub-questions (SQs) require answers:

- SQ1: What are the dominant research themes and techniques (e.g., GANs, RL) and their primary applications (e.g., level design, asset creation) in AI-driven PCG research?
- SQ2: Which countries, institutions, and publication sources are the primary centers of scientific production and impact in the field?
- SQ3: Who are the most influential and productive authors, and what are the structural patterns of the collaboration networks within the PCGML research community?

1.4 Research Scope

Four main boundaries define the scope of this study to ensure a focused, thorough and rigorous analysis:

1. Database Scope: The analysis is exclusively confined to papers available for exported from the Scopus database. Scopus was chosen for its comprehensive coverage of computer science and engineering literature, as well as its high-quality metadata for easy use in Biblioshiny.

2. Thematic Scope: The research is thematically centered on the intersection of AI and PCG in video games, specifically focusing on the modern technology of PCGML. The study deliberately excludes literature on non-ML-based PCG (e.g., purely rule-based systems) and non-generative applications of AI in gaming (e.g., traditional pathfinding or finite-state machine-based NPCs), as defined by the study's inclusion criteria and manual curation process (detailed in chapter 3).
3. Temporal Scope: The analysis covers the period from 2008 to 2025. The start year was selected to capture the earliest foundational work in the PCGML field, which began to emerge in the late 2000s, while the end date is inclusive of the most current research available at the time of data collection, ensuring the analysis reflects the contemporary state of the research field.
4. Document Scope: The dataset is limited to peer-reviewed journal articles and conference proceedings published in the English language. Other document types, such as books, book chapters, editorials, and non-English publications, are excluded from the analysis.

By defining these boundaries, this study strives to provide a deep and systematic analysis of a specific, high-impact research topic, as opposed to a general study of all types of AI in games.

1.5 Significance of the Research

As mentioned, this study endeavour to provide actionable insights for several key groups, bridging the gap between academic research and industry practice in the field of AI-driven PCGML game development. As opposed to using bibliometric analysis in order to structure a literature review, this study seek to go beyond that and build a detailed map of the PCGML landscape with a structured, data-driven understanding of the field's past, present, and future trajectory for the contribution of several segments of society:

Contribution to Academia: The study seek to provide a comprehensive bibliometric map of the PCGML research field, identify core methods, potential research gaps, intellectual turning points, collect the most influential authors in the field, key research institutions, and the primary journals and conferences for the topic. This serves as a valuable roadmap for future academic inquiry, helping researchers build an understanding of the foundational work and potentially enabling academics to identify opportunities for novel contributions and collaborations.

Contribution to Business and Industry: For developers, technical artists, and studio leaders in the video game industry, this research may offer practical and strategic insights. For this group of people, the related specific deliverables from this analysis include:

- A ranked list of the most influential academic labs and corporate research groups for potential recruitment and collaboration.
- A data-driven timeline of which generative techniques gained prominence and when, potentially aiding decisions regarding technology adoption strategies and R&D investment.
- An identification of the top journals and conferences, highlighting the best venues for professionals to track cutting-edge research and stay ahead of the curve.

Contribution to Society: While indirect, the advancement of PCGML has broader societal implications. By enabling the creation of more dynamic, personalized, and replayable interactive experiences through PCGML, it contributes to the cultural and economic impact of the digital entertainment sector. Furthermore, the generative techniques developed in gaming might find applications in other fields, such as simulation for training, architectural visualization, and film production, meaning that advancements in this niche can foster innovation more broadly.

1.6 Definition of Terms

Artificial Intelligence (AI): Within this study, AI is defined as the research field creating computational systems capable of human-like intellectual capabilities, including reasoning, learning, and autonomous action toward a specific goal. The scope of this research is narrowed specifically to the AI subfield of machine learning (Russell & Norvig, 2021).

Machine Learning (ML): As a subfield of AI, machine learning is distinct in its approach: rather than being explicitly programmed with a set of rules, ML systems are designed to learn patterns and make decisions directly from data. ML employs algorithms to analyze existing information and subsequently make predictions or choices about new unseen data (Samuel, 1959; Bishop, 2006).

Generative Adversarial Networks (GANs): GANs are a specific machine learning model built on a competitive dynamic between two neural networks. A "generator" network attempts to create new, authentic data (such as game assets), while a "discriminator" network simultaneously learns to evaluate whether that data is real or fake. This competitive training process results in the generation of novel and highly realistic content (Goodfellow et al., 2014).

Reinforcement Learning (RL): This form of machine learning involves training an intelligent "agent" to make a series of decisions within a given environment. The agent's goal is to learn a strategy (or policy) that maximizes a total reward signal over time. In the context of this research, an RL agent might be tasked with "playing" the role of a level designer, where it learns to arrange game elements in a way that creates a positive player experience, thereby earning a reward (Sutton & Barto, 2018).

Procedural Content Generation (PCG): PCG refers to the use of algorithms for the automated creation of game content, thereby reducing the need for direct human input. The scope of PCG is broad, covering everything from simple rule-based systems to complex simulations, all used to produce vast quantities of unique game elements like levels, maps, or items (Short & Adams, 2017).

Procedural Content Generation via Machine Learning (PCGML): As the primary focus of this study, PCGML represents an evolution of traditional PCG. Its key distinction is the explicit use of ML models (such as neural networks or reinforcement learning agents). These models are trained on existing game data to learn complex patterns, which allows them to generate new content that is not only novel but also stylistically consistent with the training data (Summerville et al., 2018).

Bibliometrics: The use of statistical and quantitative methods to analyze academic literature, such as articles, and other publications. This methodology is used to map: intellectual structure of a research field, identify influential works, and to uncover trends and patterns of collaboration within it. (Pritchard, 1969; Aria & Cuccurullo, 2017).

CHAPTER 2
LITERATURE REVIEW

This chapter provides a rigorous explanation of the theoretical and methodological topics for the study at hand. This chapter's function differs from that of a standard literature review. Instead of presenting an exhaustive summary of PCGML research, its purpose is twofold and specifically designed to support the bibliometric methodology: Firstly, it will review the literature on bibliometrics as a research methodology, outlining its principles and justifying its application for mapping a scientific field; Secondly, key foundational concepts and essential academia with regards to PCGML are discussed, with the intent to establish the core theoretical context for the subsequent analysis.

2.1 Bibliometrics as a Method for Science Mapping

To understand the structure and evolution of a research domain, scholars may employ bibliometric analysis, a quantitative method that uses statistical data from academic publications, such as articles and citations (Pritchard, 1969). This approach moves beyond subjective qualitative reviews by seeking to build an objective, data-driven "map" of a field's intellectual landscape. This is done by systematically analyzing large-scale patterns within publication data, which in turn, makes it possible to identify a field's central research themes, pinpoint its most influential authors and institutions, and chart the collaborative networks that are vital for creating knowledge (Aria & Cuccurullo, 2017).

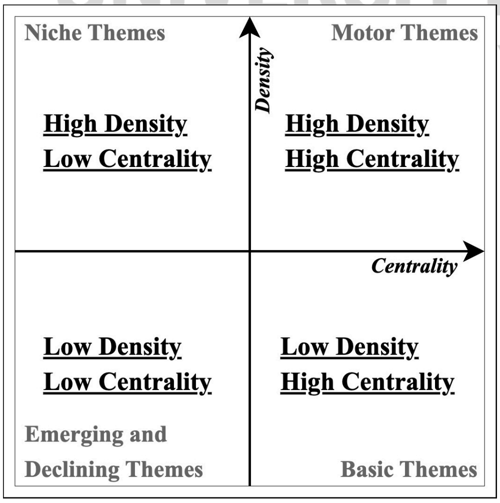
This quantitative method proves especially valuable for a field like PCGML, which is characterized by rapid growth and a high degree of fragmentation. For fragmented fields, a traditional literature review is often impractical; bibliometrics overcomes this by providing a macroscopic, data-driven overview that can structure a large and scattered body of literature, identifying its core concepts and intellectual pillars without succumbing to researcher bias (Donthu et al., 2021).

Given the large volume of work, a traditional manual literature review would be a challenging task. Bibliometrics offers the necessary tools to process thousands of publications, allowing for the identification of core research areas. Furthermore, through citation analysis, the scholarly impact of individual works can be analyzed, while co-authorship analysis helps visualize the social structure of the research community itself (Donthu, Kumar, Mukherjee, Pandey, & Lim, 2021). The final result is a reproducible overview that builds a solid foundation for understanding how the field has developed.

2.2 Bibliometrics: Thematic Mapping

Beyond mapping the social networks of authors and institutions, bibliometric analysis can be used to visualize the conceptual structure of a research field, revealing how key ideas and themes are organized and interconnected. One of the most powerful techniques for this is thematic mapping, a method developed by Cobo, López-Herrera, Herrera-Viedma, and Herrera (2011). This approach uses a co-word analysis to cluster keywords into themes and then plots them on a two-dimensional strategic diagram (example in figure 2.1).

Figure 2.1: The Strategic Diagram for Thematic Mapping.



Source: Adapted from Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. Journal of Informetrics, 5(1), 148-166.

A thematic map (figure 2.1) is defined by two key axes: Centrality and Density. Understanding these dimensions is crucial for interpreting the map's strategic implications:

- Centrality (Relevance): This axis measures the degree of interaction of a keyword cluster with other clusters in the research field being amalyzed. In simpler terms, it indicates a theme's importance and its role in connecting different research topics, and is relevant to many other areas of the particular segment of research.
- Density (Development): This axis measures the internal strength of the links among the keywords that constitute a single theme. A theme with high density is a well-established research program with a strong, self-contained conceptual framework where the inherent ideas are tightly linked.

As can be seen from figure 2.1, the thematic map is divided into four quadrants, each representing a different type of theme with a distinct role in the research field being analyzed (Cobo et al., 2011; Aria & Cuccurullo, 2017).

5. Motor Themes (Top-Right Quadrant): These themes are considered the "engine" of the research field. They are both well-developed (high density) and crucial to the entire domain (high centrality). These are mature, foundational topics that drive progress and are closely connected to other parts of the research field.
6. Niche Themes (Top-Left Quadrant): These are highly developed, specialized themes that have marginal importance to the broader field. They represent more isolated groups of experts in a specific sub-field. While internally coherent and mature (high density), they are not well-connected to other themes.
7. Basic Themes (Bottom-Right Quadrant): These themes are important to the field but are not well-developed. Their high centrality suggests they are concepts that cut across many research areas, but their low density suggests a lack of a focused, mature research program. These often represent general, fundamental or overarching topics. For example, in a study of "Artificial Intelligence," the term *ethics* would

likely be a basic theme. It is highly central because it is relevant to nearly every AI application, however it is not densely developed because the ethical research in different AI sub-fields (e.g., autonomous vehicles versus medical AI) are distinct and do not form a single, coherent research cluster.

8. Emerging or Declining Themes: These themes are both weak in current development less important. This quadrant is particularly dynamic; it can contain themes that are just beginning to emerge and have not yet built up a research community (emerging), or themes that are becoming obsolete or have been abandoned (declining). A temporal analysis is used to distinguish between these two possibilities.

This analytical framework, grounded in the work of Cobo et al. (2011) and implemented through the bibliometrix R package (Aria & Cuccurullo, 2017), will be applied in Chapter 4 to analyze the structure of the PCGML research field and interpret the strategic roles of its dominant themes.

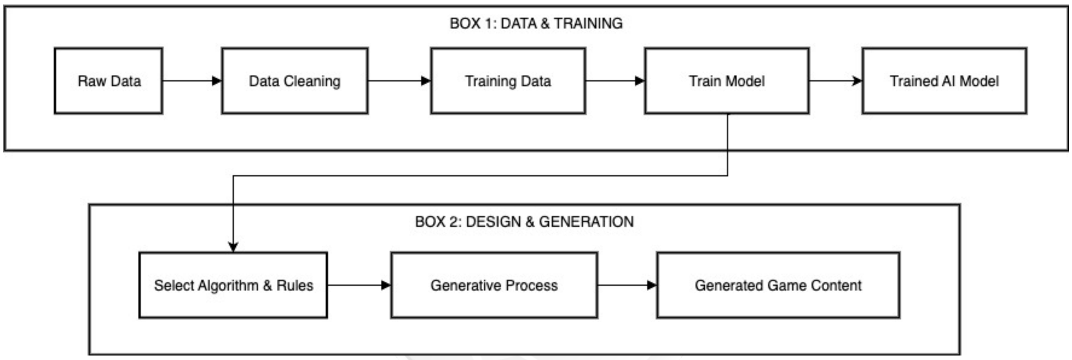
2.3 Foundational Concepts in AI-Driven Procedural Content Generation

A critical paper from Summerville et al. (2018) was instrumental in formally defining the PCGML research field. The authors categorized the different ways machine learning could be applied to generative tasks, including using supervised learning to copy existing styles, reinforcement learning to optimize for a goal like player enjoyment, or unsupervised learning to find novel patterns in game data.

Several key practical studies have produced this theoretical framework. For example, the work of Volz et al. (2018) showed the power of Generative Adversarial Networks (GANs) in generating playable levels for the classic game Super Mario Bros. Likewise, the "Experience-Driven PCG" concept, which seeks to create content that directly influences a player's emotional or cognitive state, has been a significant theoretical contribution to the field (Yannakakis & Liapis, 2017).

Collectively, these foundational works provide the conceptual basis for the large-scale analysis of this study. The foundational work establishes the key techniques and goals that one would expect to be reflected in the broader bibliometric data. The general end-to-end workflow described in the previously mentioned literature, from initial data processing to the final generated output, is visualized in Figure 2.2:

Figure 2.2: The PCGML Workflow.

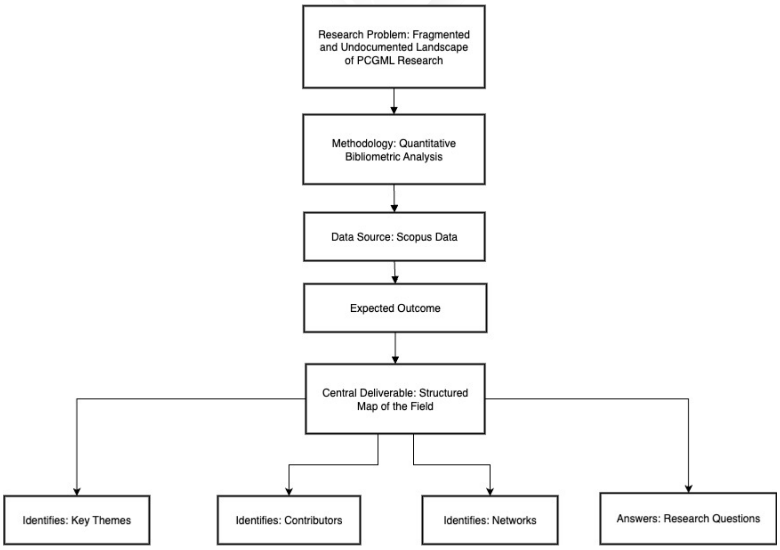


Source: Author's elaboration, based on Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A. K., Isaksen, A., ... & Togelius, J. (2018). Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games,* 10(3), 257–270.

CHAPTER 3
RESEARCH DESIGN AND METHODOLOGY

This chapter describes the specific research design and methodology employed to conduct the bibliometric analysis of the PCGML field. Building on the theoretical foundations of bibliometrics outlined in Chapter 2, this section describes the practical steps taken for data collection, pre-processing, and analysis. The methodology follows the PRISMA guidelines for transparency (Page et al., 2021) and utilizes the bibliometrix R package. The overall research workflow is visualized in Figure 3.1 below:

Figure 3.1: Research Design Flowchart.


Research Problem: Fragmented and Undocumented Landscape of PCGML Research
Methodology: Quantitative Bibliometric Analysis
Data Source: Scopus Data
Expected Outcome
Central Deliverable: Structured Map of the Field
Identifies: Key Themes
Identifies: Contributors
Identifies: Networks
Answers: Research Questions

Source: Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* 372.

3.1 Research Setup

This study employs a quantitative and descriptive bibliometric analysis to achieve its objectives. The research has been designed specifically to map the intellectual and social landscape of the PCGML field systematically. To maintain objectivity, the process avoids any qualitative interpretation of the source documents' content; instead, the entire analytical workflow is carried out in a computational environment with specialized software.

For the primary analysis, this research utilizes the bibliometrix R package, an open-source toolset created for comprehensive science mapping and scientometric analysis (Aria & Cuccurullo, 2017). The interactive web-based interface, Biblioshiny, was used for the hands-on analysis. This platform was selected because its robust features for data import, cleaning, network analysis, and visualization directly align with the needs of this study's research questions. To ensure the methodology is both transparent and reproducible, the overall research setup follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, guiding the process from the initial data collection to the final reporting of results (Page et al., 2021).

3.2 Sampling and Data Collection

To compile the dataset for this analysis, the Scopus database was chosen as the primary source. The decision to use Scopus was based on several key factors: its comprehensive coverage of peer-reviewed literature in computer science and engineering, the high quality of its structured metadata, and its powerful citation tracking features. These characteristics are all essential for conducting a rigorous bibliometric study.

To ensure the compiled data was both precise and relevant, a multi-layered Boolean search query was developed over an iterative testing period. The goal of this query was to effectively isolate literature focused on Procedural Content Generation via Machine Learning (PCGML). This final query was applied to the titles, abstracts, and keywords of documents published in the period from 2008 to 2025. The 2008 start date was selected to capture the field's earliest foundational papers, while the 2025 end date allows for the inclusion of early-access publications, ensuring the analysis is as current as possible. The final search query submitted to the Scopus database was:

TITLE-ABS-KEY (("AI" OR "artificial intelligence" OR "Artificial Agent" OR "machine learning" OR "deep learning" OR "neural networks" OR "generative AI") AND ("video game" OR "gaming" OR "game development" OR "Algorithmic Content Creation" OR "Dynamic Content Generation" OR "Automatic Game

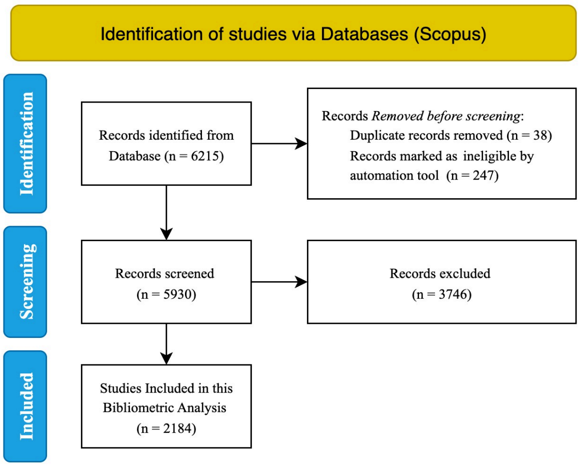
Design" OR "Generative Design in Games" OR "Procedural Level Generation" OR "Automated Content Generation" OR "Random Content Generation" OR "Computational Creativity in Gaming" OR "experience-driven pcg" OR "search-based pcg")) AND (LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "k") OR LIMIT-TO (SRCTYPE , "b")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "bk")) AND (LIMIT-TO (LANGUAGE , "English"))

This query structure establishes the context for AI and PCGML while filtering for specific machine learning applications. The execution of this query and subsequent filtering for document type (article, proceeding) and language (English) yielded an initially large dataset of 6,215 documents for analysis. As suspected, there were quite alot of unrelated papers to remove from the dataset, with the process of that discussed in the next section of the paper.

3.3 Data Curation and Screening

The initial Scopus search returned a large pool of 6,215 documents. To distill this broad collection into a thematically relevant dataset, a careful, multi-stage curation was necessary. This entire filtering process, which is visually summarized in the PRISMA diagram (Figure 3.2), was essential for the integrity of the analysis:

Figure 3.2: PRISMA Flowchart of the Study Selection Process


Identification of studies via Databases (Scopus)
Identification
Screening
Included
Records identified from Database (n = 6215)
Records *Removed before screening*:
Duplicate records removed (n = 38)
Records marked as ineligible by automation tool (n = 247)
Records screened (n = 5930)
Records excluded (n = 3746)
Studies Included in this Bibliometric Analysis (n = 2184)

Source: Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* 372.

The first pass of cleaning the data was automated through the use of Rayyan. This initial step removed 38 duplicate records and an additional 247 records that were flagged as ineligible during the import process due to being from books and book chapters, as well as being from 2007 and earlier (the focus of this paper is 2008-2025). This left 5,930 unique articles and conference papers requiring further review and screening.

The core of the curation work involved screening the titles and abstracts of these thousands of papers against a strict, predefined set of rules, which was crucial for ensuring that every paper in the final set was directly relevant to PCGML. To be retained or discarded for the final analysis, a paper had to focus primarily on the following:

Table 3.1: Overview of Inclusion and Exclusion Criteria for Manual Screening

Inclusion (paper primarily focused on)	Exclusion (paper primarily focused on)
1. The generation of game levels, maps, rules, or environments. 2. The generation of game assets (e.g., models, textures, animations). 3. The generation of narrative elements (e.g., quests, dialogue, plots). 4. It is a theoretical or survey paper that explicitly analyzes the PCGML field.	1. Non-generative AI tasks like NPC pathfinding or player modeling. 2. Using games simply as a tool to study unrelated fields, such as medicine. 3. Unrelated computer science theory without a direct, applied link to PCGML. 4. Being a data artifact (e.g., an erratum, preface, or non-English publication).

Applying these criteria resulted in the exclusion of 3,746 documents, resulting in a final, focused dataset containing 2,184 papers. This refined collection forms the foundation for the entire bibliometric analysis that follows.

3.4 Data Analysis Strategy

As mentioned earlier, this study employs a quantitative research approach, utilizing a suite of bibliometric analysis techniques to map the PCGML research field systematically. All analyses were conducted using the Biblioshiny for R software package. The relationship between the research questions, the chosen analytical methods, and the specific functions used is summarized in Table 3.2 and detailed below:

Table 3.2: Analytical Strategy and Methods Regarding Research Questions

Research Questions	Bibliometric Analysis Method	Biblioshiny Function(s)
RQ1: What are the primary techniques and applications within AI-driven PCG research?	- Thematic analysis - Trend analysis	- Thematic Map - Trend Topics - Word Cloud
RQ2: What are the primary regional and institutional hubs of expertise?	- Geographical analysis - Institutional analysis	- Country Scientific Production - Country Collaboration Map - Most Relevant Affiliations
RQ3: Who are the most influential authors, and what do their collaboration networks look like?	- Productivity analysis - Citation analysis - Co-authorship analysis	- Most Productive/Cited Authors - Most Relevant Sources - Co-Authorship Network

To answer RQ1 a conceptual structure analysis was performed. This involved generating a Thematic Map based on keyword co-occurrence in abstracts to identify and categorize research themes. To complement the thematic map, a Trend Topics analysis was also used to chart the chronological development of these themes, revealing which topics gained or lost prominence over the years.

The second research question (RQ2), concerning the field's primary hubs of expertise, was addressed through a performance and source analysis. This method systematically identifies the most significant contributors by ranking countries, institutions, and publication venues (i.e., journals and conferences) based on both their total scientific output and their cumulative citation impact.

Finally, to investigate the influential authors and networks central to the third research question (RQ3), the study performed a social structure analysis. This involved first pinpointing the most productive and most cited authors. Subsequently, a co-authorship network graph was generated to create a visual map of the collaborative relationships and distinct research clusters that form the PCGML research community.

CHAPTER 4
FINDINGS

This chapter details the findings from the conducted bibliometric analysis of the curated dataset, as described in the previous chapter. The results are organized according to the three SQs presented in chapter 2, and starts with a high-level statistical profile of the data, then moves to a deep dive into the field's dominant themes and their evolution (SQ1), and subsequently, it identifies the key centres of scientific production and impact (SQ2) before concluding with an analysis of the most influential authors and their collaboration networks (SQ3).

4.1 A Statistical Profile of the Research Corpus

A high-level statistical analysis of the data reveals the foundational characteristics and the key metrics of the PCGML research field between 2008 and 2025, which comprises 2,184 documents from 989 unique sources, are presented in Table 4.1 below:

Table 4.1: Key Metrics of the Dataset

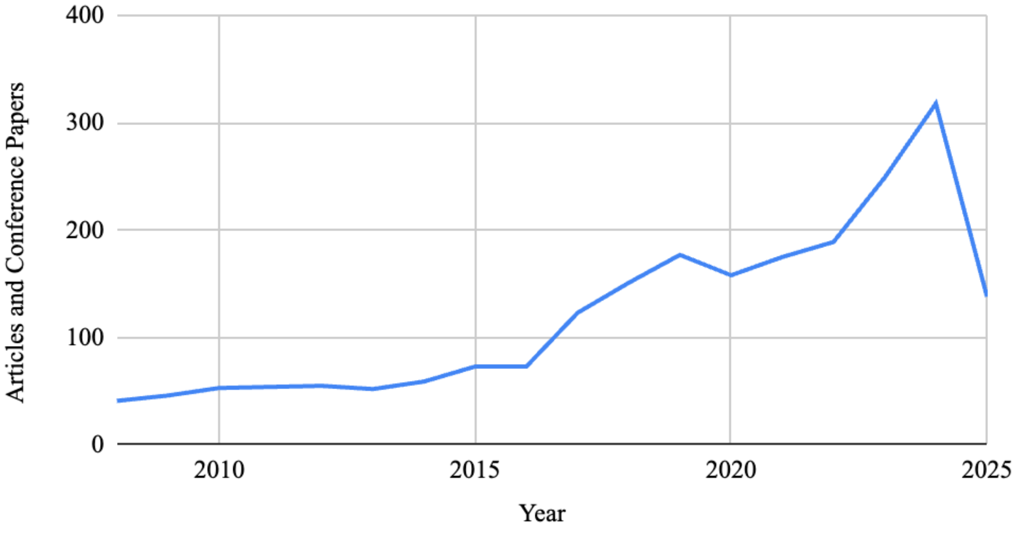
Description	***Results***
Timespan	2008:2025
Sources	989
Documents	2,184
Annual Growth Rate	7.4
Avg. citations per doc	2.49
Co-Authors per doc	3.64
Int co-authorships %	16.99
Articles	503
Conference papers	1,681

Between 2008 and 2025, the research demonstrated a healthy Annual Growth Rate of 7.4%, growing academic interest. The scholarly impact is notable, with an average of 2.49 citations per document, suggesting that research in this domain is actively being engaged with and built upon.

In addition, the generated academic material are overwhelmingly conference-driven: Conference papers (1,681) outnumber journal articles (503) by a ratio of approximately 3.3-to-1. This is typical of computer science, which is widely recognized as a discipline where high-impact research is primarily spread through peer-reviewed conferences rather than journals (Franceschet, 2010).

The overall global growth of the field over the specified timespan is visualized in Figure 4.1 below:

Figure 4.1: Annual Publication Volume in PCGML Research.



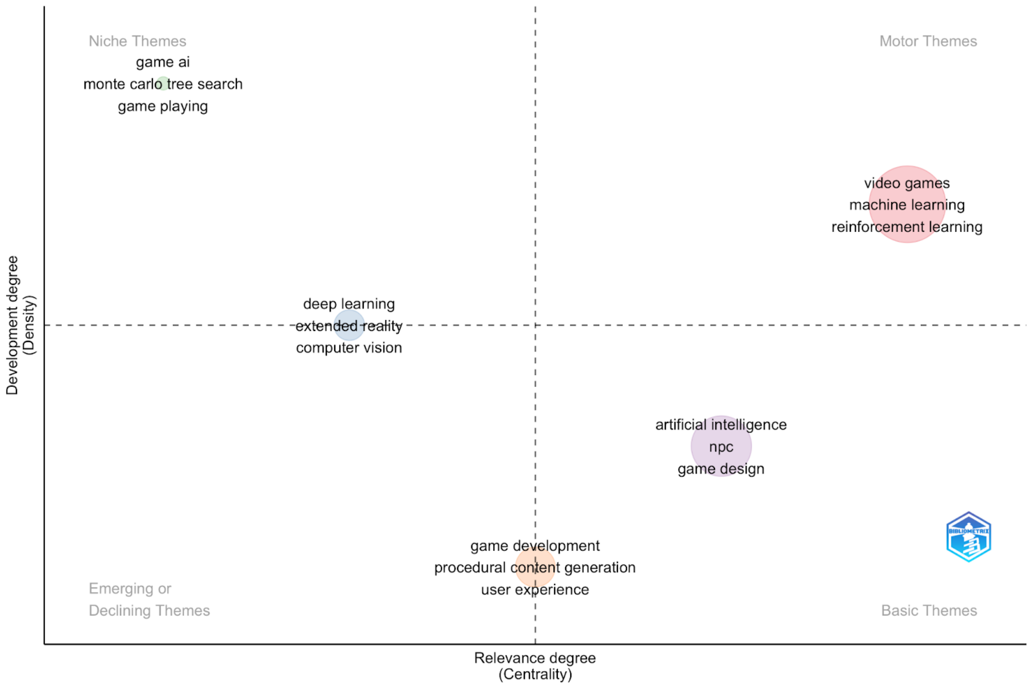
The figure reveals several distinct phases in the research field's development. 2008 to 2014, the field was in an early stage, characterized by low and relatively flat. A clear inflection point occurred around 2014, kicking off a period of sustained and rapid growth that continued until 2019.

A notable dip in productivity is observed in 2020, which may reflect a disruption in research activities due to the global COVID-19 pandemic; nonetheless, the field seems to have strong resilience, with output quickly recovering and accelerating from 2021 to 2024. This period culminates in an large peak in 2024, with over 300 articles published, indicating a surge of interest likely fueled by the broader boom in generative AI. The sharp drop-off in 2025 is a data artifact, reflecting the incomplete collection of early-access articles for the upcoming year, and does not signify a decline in the field.

4.2 Answering SQ1: Thematic Landscape and Conceptual Evolution

To answer the first research question regarding dominant themes and their evolution, a thematic map and a trend topics analysis were conducted. The thematic map for the study is presented below:

Figure 4.2: Thematic Map of the PCGML research field.



The thematic map above (Figure 4.2) provides a clear visualization of the PCGML research field's conceptual structure. As mentioned in earlier chapters, plotting research themes on two axes defined as: Centrality (the degree in which it interacts with other themes, indicating its relevance to the overall field) and Density (the strength of the internal links between keywords within a theme, indicating its level of development) the map categorizes themes into four distinct quadrants. Each quadrant reveals a different role that a research theme plays within the intellectual landscape.

4.2.1 Motor Themes (Top-Right): High Density and High Centrality

The top-right quadrant is reserved for themes that are both highly central and highly dense. In bibliometric theory, these are considered the "motor" themes because they are well-developed, fields of research that are foundational and commonly drive progress across the entire field. In this analysis, the Motor quadrant is occupied by a single well-developed cluster comprising machine learning, reinforcement learning, and neural network. The interpretation of this placement is critical:

- High Centrality: The centrality of this cluster indicates that these topics function as major intellectual hubs. Nearly every paper across all other quadrants, whether focused on game design, niche algorithms, or emerging applications, must in some way engage with or build upon these core ML concepts. They are the fundamental language of modern PCGML research.
- High Density: The high density reveals that machine learning, reinforcement learning, and neural networks are not loosely related terms. They form a tight, internally coherent research program where advancements build directly upon one another, creating a well-developed and specialized area of technology.

The combination of these two properties makes this a motor theme which means that it is not a passive background topic, it is an active and mature area of research that produces essential knowledge and technology that power the rest of the field. This finding suggests that the primary driver of innovation in modern PCGML is fundamentally methodological, as all three of the components work together to allow for PCGML to even be possible; these three are what will enable the research field to flourish and push boundaries.

4.2.2 Basic Themes (Bottom-Right): The Core Domain

This quadrant is characterized by high centrality but low density. These themes are important and flexible, connecting to many other topics, but they do not form a highly developed, specialized research program on their own. The map places the large cluster of video games, artificial intelligence, and game design here.

- High Centrality: The high centrality is expected, as these terms define the entire research domain. For example, a paper on PCGML is inherently a paper about *artificial intelligence* and *video games*, making these themes relevant to nearly every document in the dataset.
- Low Density: The low density is equally revealing. It signifies that while nearly every paper in the dataset is about "AI in video games," the novel research contributions are not focused on defining what a video game is. Instead, the innovative work is detailed within the more specialized parts, namely in the Motor and Niche themes.

In conclusion, this cluster represents the foundation that Procedural Content Generation via Machine Learning leans on when making use of machine learning, reinforcement learning, and neural network application in the research field.

4.2.3 Niche Themes (Top-Left): The Specialized Toolkits

Niche themes exhibit high density but lower centrality. They represent well-developed, specialized topics that are important to a specific sub-community but are not as globally relevant as motor or basic themes. This analysis identifies the cluster containing monte carlo tree search, game playing, and the gvg-ai competition as a key niche.

- High Density: The high density indicates a mature and self-contained research field. The researchers working on search-based AI for competitive game-playing seemingly form a tight-knit community. Their papers reference a shared body of knowledge and techniques, creating strong internal links between these keywords.
- Low Centrality: The lower centrality correctly positions it as a specialized toolkit. While crucial for specific problems like creating super-human game-playing agents, Monte Carlo Tree Search is not a technique applied in every PCGML paper, unlike the more general methods found in the Motor Themes. A paper on generating art assets with GANs, for example, would likely not engage with this theme.

This quadrant thus highlights the existence of expert sub-fields within PCGML. These are areas with deep knowledge and mature methodologies that solve specific, important problems, even if those solutions are not universally applicable across the entire domain.

4.2.4 Emerging or Declining Themes (Bottom-Left)

This quadrant, defined by low centrality and low density, represents themes that are either new and not yet fully integrated into the field, or older topics that are losing relevance. The Trend Topics plot (Figure 4.3) helps distinguish between the two. In this case, the themes are clearly emerging application frontiers.

- Low Density & Low Centrality: This cluster, containing extended reality, computer vision, and deep learning, represents a modern research frontier. Coupling the placement of the cluster with the later discussed details of the Trend Topics Plot, the data seem to indicate that these topics are new to the PCGML research field. The research communities are still forming, and the connections between, for instance, extended reality (VR/AR) and core PCGML methods are not yet as established or dense as those within the Motor Themes.

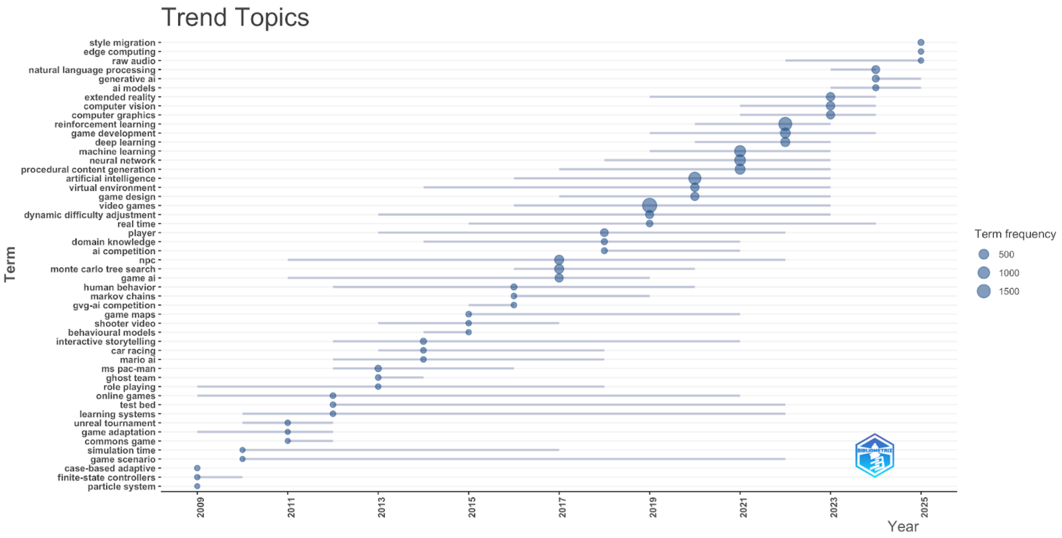
4.2.5 Thematic Map: Game development, PCG, UX

The placement of the *game development, procedural content generation, and user experience* cluster is particularly interesting. It is not a declining theme but rather represents an application context and the goal of the research studied. Its position suggests that while central to the field's purpose, the novel scientific contributions tracked by bibliometrics are more frequently found in the methods (Motor Themes) than in the direct, empirical study of the development process or user experience itself.

4.2.6 Trend Topics of PCGML Research Field

While the thematic map provides a static snapshot of the the research fields structure, the Trend Topics plot (Figure 4.3) offers a longitudinal, chronological representation. By tracking the prominence of keywords over time, the plot reveals a clear technological journey marked by three distinct time periods.

Figure 4.3: Longitudinal plot of Trend Topics in PCGML research.



Period 1: The Age of Classic AI (2008-2015)

The lower portion of the plot, representing the earliest years of the dataset, tells the story of the field's foundations in classic AI. This period is characterized by the presence of terms like monte carlo tree search, markov chains, and game-specific agents such as mario ai and ms pac-man. The focus during this era was on solving well-defined problems using algorithms and statistical methods with the goal to create intelligent behavior through explicit logic (like finite-state controllers, a term seen in the raw data) or sophisticated search trees (like MCTS). This was a "problem-solving" period, where AI was a tool for creating an opponent or optimizing a specific game mechanic. The research seem to often be tied to specific games or competitions, offering practical solutions.

Period 2: The Machine Learning Revolution (2016-2021)

The middle of the plot shows a dramatic shift, indicating a widespread adoption of modern machine learning. This time period is defined by a substantial growth in frequency and prominence of the "Motor Themes" identified earlier: machine learning, neural networks, and reinforcement learning. The dots for these terms become significantly larger, indicating they are dominating the research conversation.

This marks a fundamental shift away from programming explicit rules to creating systems that learn from data. Made possible by evident breakthroughs in technology, researchers in PCGML began applying deep learning and deep reinforcement learning at scale.

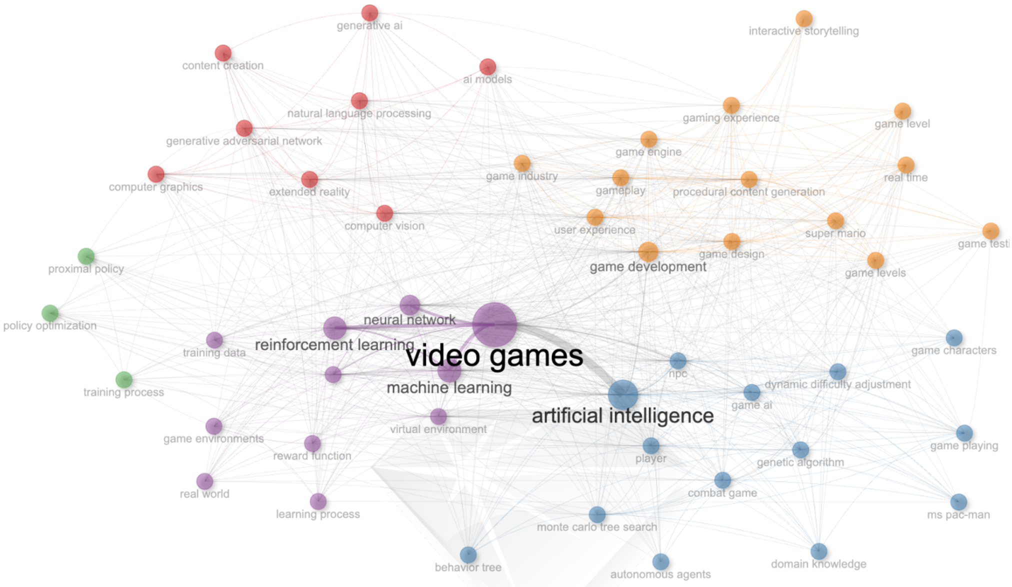
Period 3: The Deep Generative Frontier (2022-Present)

The top of the plot reveals the most current state of the research field, characterized by a move toward highly sophisticated, large-scale generative models. The most recent years show an emergence of new keywords like *generative ai, ai models* (implying large, pre-trained models like Transformers), *natural language processing, computer vision,* and *extended reality*. The rise of generative AI ushers in a focus on creating entire pieces of content: art, music, narrative, and dialogue (natural language processing), at a quality and scale previously unimaginable. The emergence of computer vision and extended reality indicates that these powerful new models are further being developed for the next generation of 3D and immersive gaming platforms. This is the new direction the established methods of the Machine Learning Revolution are being pushed to new creative and technical limits.

In summary, the Trend Topics plot visualizes a clear technological journey: from classic, rule-based systems, through a revolutionary shift to data-driven machine learning, to the current state of sophisticated, deep generative models. This evolutionary path provides essential context for understanding the field's current state and future direction.

Finally, to provide a detailed view of the conceptual relationships, a keyword co-occurrence network was generated (Figure 4.4). This network visualizes which terms frequently appear together in the same documents, effectively mapping the intellectual connections that form the field's structure. The map confirms and further elaborates on the findings from the thematic and trend analyses, revealing several distinct research clusters linked by central hubs.

Figure 4.4: Co-occurrence Network of Keywords in PCGML Research.



The network is anchored by two dominant central nodes: artificial intelligence and video games. These act as the main nodes connecting the distinct different clusters, reinforcing their role as the foundational context for the entire field. The specific clusters orbiting these hubs each tell a story of a particular research focus or methodology.

The Green Cluster: The Reinforcement Learning Workflow.

This prominent cluster illustrates the complete workflow of modern reinforcement learning research in gaming: reinforcement learning, training data, reward function, policy optimization (and its specific algorithm, proximal policy), and training process all describe the process of PCGML. Researchers start with training data from a game environment, design a reward function to guide the learning, and then engage in a training process using a policy optimization algorithm to produce an intelligent agent. This cluster represents the practical, step-by-step methodology of a significant portion of the PCGML research community.

The Red Cluster: The Game Design and Application Domain

This cluster represents the core problem domain and the ultimate goal of the research: the application of AI within game development. It connects the high-level goals of game design and game development with the tangible outputs, such as creating game levels or other content (procedural content generation). The aim is to influence the gameplay and improve the overall user experience directly. This cluster represents the practical challenges and objectives that the technical methods from the other clusters are trying to solve.

The Blue Cluster: The Modern Deep Generative Toolkit

This cluster, closely connected to the "Generative Frontier" era from the trend analysis, represents the cutting-edge tools for content creation. The narrative here is about direct content synthesis using sophisticated models. Researchers use powerful AI models like generative adversarial networks for content creation. The inclusion of natural language processing shows this extends beyond visual assets to text and dialogue. The central term, generative ai, acts as the umbrella for this modern toolkit, signifying a move from AI as a player-agent to AI as a content-creator.

The Purple Cluster: The Classic Agent-Based AI

This cluster tells the story of the field's in creating autonomous, game-playing agents using classic AI techniques. The goal is to create autonomous agents that excel at game playing in specific testbeds like ms pac-man. The tools for this are established algorithms like *monte carlo tree search* and *genetic algorithms*. This cluster represents a mature, foundational pillar of AI in games, focused on agent behavior rather than content generation, but it provides the intellectual heritage for much of the agent-based work still done today.

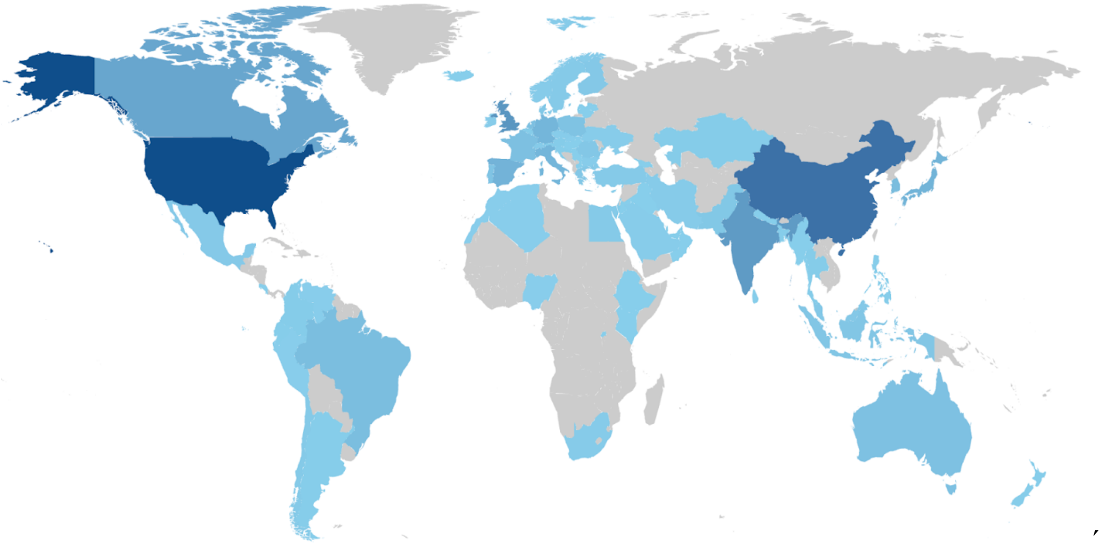
4.3 Answering SQ2: Hubs of Scientific Production and Impact

This section addresses the second research question by identifying the leading countries, institutions, and publication sources in PCGML research. The analysis moves from a broad geographical overview to the specific institutional players and publication venues, revealing a complex landscape where research volume and scholarly impact tell different stories.

Table 4.2: Top 15 Scientific Production and Citation

Country	Produced	Citated	Avg. Citations
USA	1453	2175	1.50
China	1025	976	0.95
UK	622	8486	13.64
India	563	336	0.60
Canada	435	260	0.60
Spain	301	284	0.94
Japan	292	1407	4.82
Italy	269	272	1.01
Germany	263	346	1.32
South Korea	185	286	1.55
Brazil	175	62	0.35
Netherlands	150	509	3.39
France	145	107	0.74
Australia	133	236	1.77
Poland	113	116	1.03

Figure 4.5: Geographical Distribution of PCGML Publications.



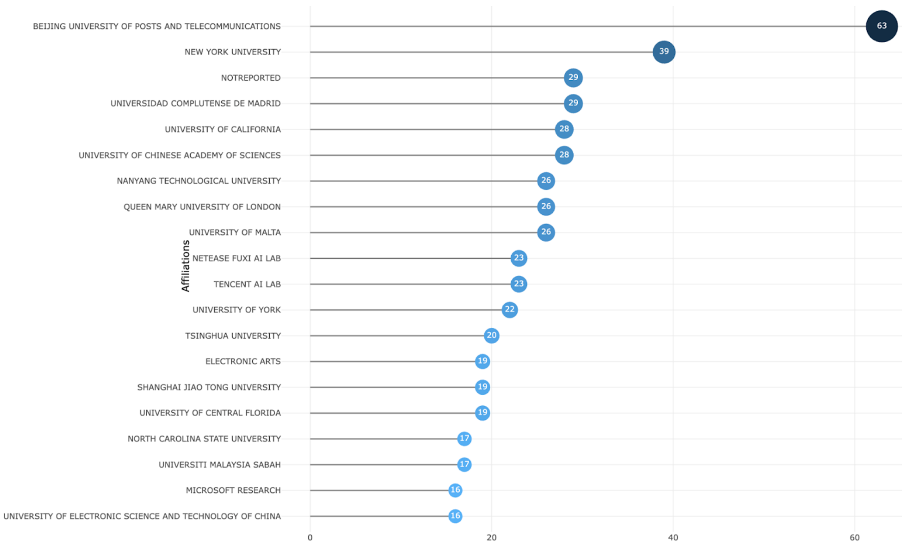
At the national level, the data reveals a critical distinction between production volume and scholarly influence. As shown in Table 4.2 and visualized in Figure 4.5, the United States and China are the undisputed leaders in the sheer quantity of research produced. This high volume reflects a broad and active research in both countries, a wide range of studies from large to smaller, more niche investigations.

However, the most striking insight emerges from the citation data. The United Kingdom, while producing significantly fewer papers, demonstrates a disproportionately high scholarly impact, leading all nations with an average of 13.64 citations per document. This is nearly an order of magnitude higher than most other countries on the list. This disparity suggests a potential difference in research strategy with the UK's output may be more concentrated on foundational, theoretical, or high-risk, high-reward studies that, when successful, become cornerstone references for the entire field. In contrast, the high volume from the US and China may represent a more diverse portfolio of work, including more incremental and applied studies that, while valuable, garner fewer citations individually.

4.3.1 The Institutional Ecosystem

The analysis of institutional contributors (Table 4.6) identifies the specific universities and corporate labs driving the field forward:

Figure 4.6: The Most Productive Affiliations in PCGML Research.



The institutional data hints at an interesting story: the PCGML field is a blended ecosystem, not a simple academic-to-industry handoff.

- Dual Research Hubs: The field is not dominated by universities. Academic powerhouses are ranked alongside well-funded corporate labs from companies like Tencent, Netease, and Microsoft, with both contributing a significant number of publications.
- Beyond Theory vs. Application: The data suggests that both sectors are conducting primary research. Corporate labs leverage unique advantages in data and computing power to tackle large-scale problems, while universities remain central to training and exploratory research.
- High Commercial Stakes: The strong presence of major tech and gaming companies is direct evidence of the field's immediate commercial value. It points to a rapid cycle where research breakthroughs are quickly developed into practical applications.

4.3.2 Publication Venues: Mapping the Flow of Knowledge

Figure 4.7: Most Relevant Sources by Publication Volume.

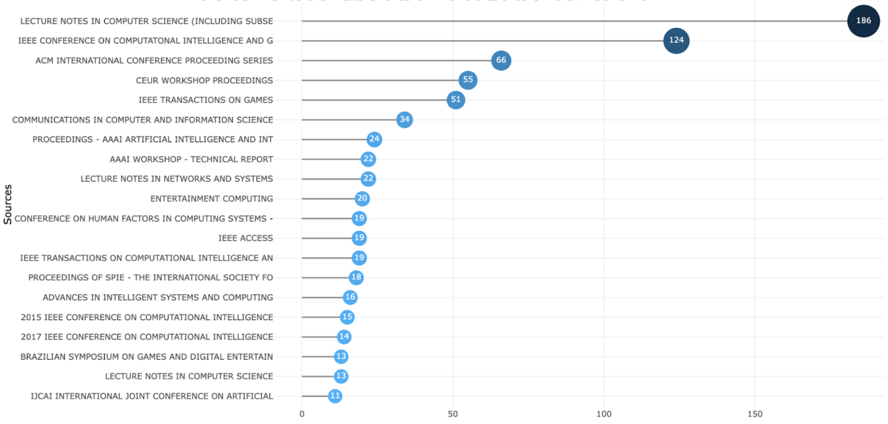


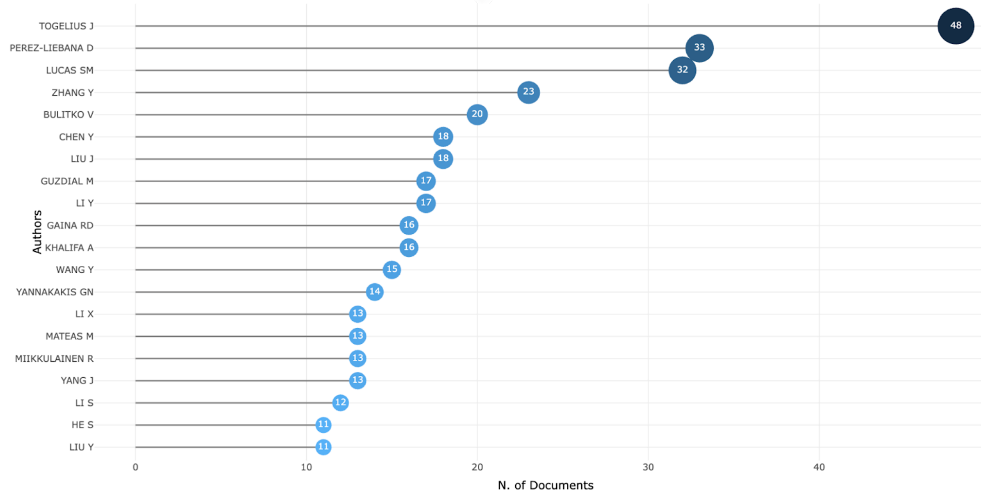
Figure 4.7 answers a critical question for anyone that is actively working or interested in the field: where do the most important research happen? The answer is overwhelmingly at conferences.

The dominance of outlets like Lecture Notes in Computer Science and proceedings from IEEE and ACM indicates that the field prioritizes the rapid spreading and learning of new ideas as opposed to slower journal timelines, is a hallmark of a field where the technology evolves quickly. For researchers and practitioners alike, monitoring these specific venues is the most effective strategy for staying on the cutting edge and anticipating future trends.

4.4 Answering SQ3: Key Authors and Collaboration Networks

The analysis of author productivity reveals a highly concentrated field. As shown in Figure 4.8, the research output is not evenly distributed; Julian Togelius emerges as the most productive author with 48 publications.

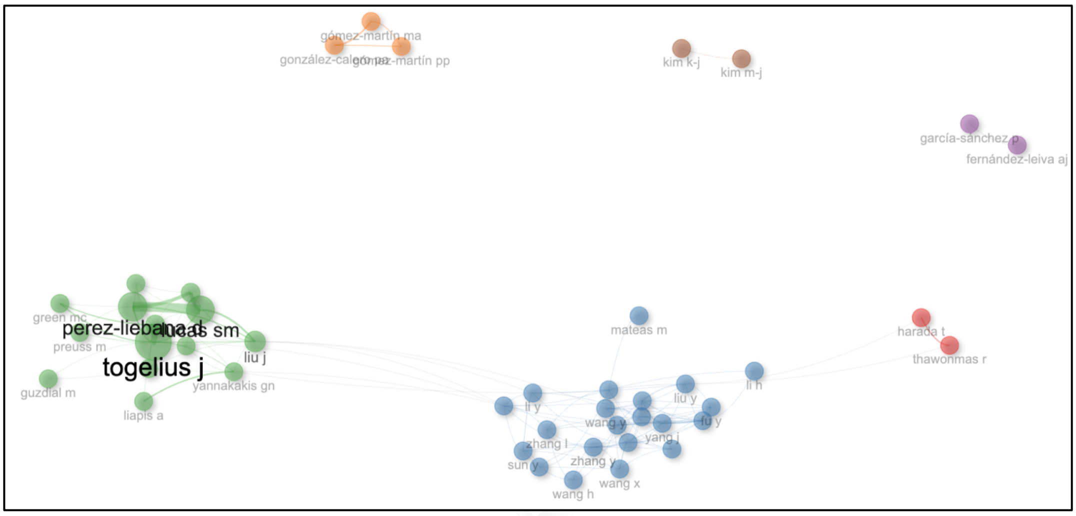
Figure 4.8: Top 20 Most Productive Authors in PCGML Research.



Additionally, a more significant finding lies in the composition of the top-ranked authors. The strong presence of Togelius's frequent collaborators, including Perez-Liebana D., Lucas SM, Guzdial M., and Yannakakis GN, points to more than just individual achievement, which indicates that there is a central research group publishing a substantial portion of the field's published work, and potentially the core research agenda is driven by this single, highly interconnected network.

4.4.1 Collaboration Network of PCGML Authors

Figure 4.9: Collaboration Network of PCGML Authors.



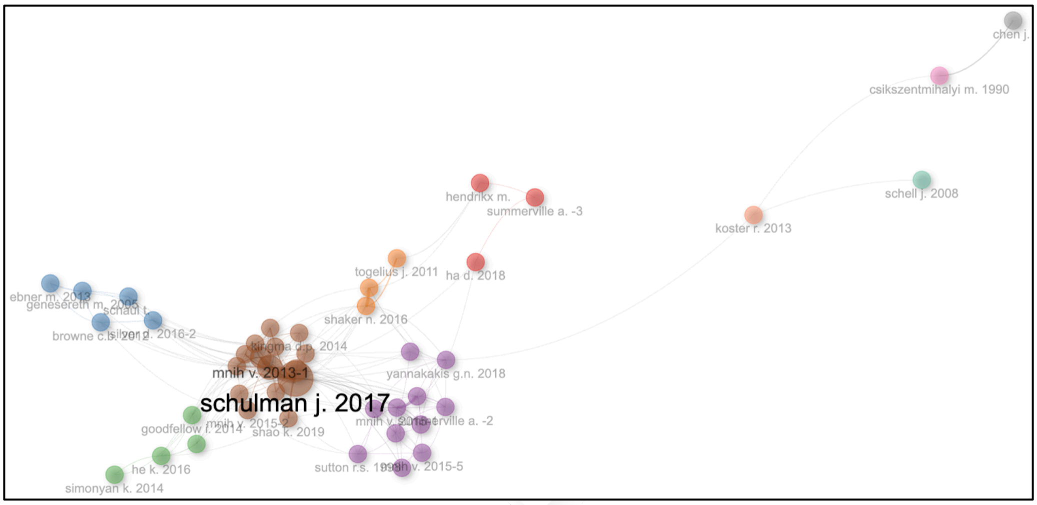
The co-authorship network in Figure 4.9 visualizes the field's social structure, which seems to indicate that the field is separated into two main components:

- A Core Cluster: A large, dense green group, organized around the most productive authors, serves as the field's center of gravity. Its high density indicates a deeply integrated and influential research program driving the main agenda.
- Smaller Clusters: Several smaller, unconnected clusters represent independent research groups. These groups might pursue more specialized topics.

Together, this structure might foster both focused progress within the central cluster and more exploratory research in the smaller ones.

4.4.2 Intellectual Structure: Uncovering the Foundational Pillars

Figure 4.10 Co-citation Network of Influential Works in PCGML Research.



For the last analysis the field's intellectual foundations, namely the co-citation network in Figure 4.10 that maps its most influential works has been examined. The analysis indicates that the intellectual structure of PCGML is built upon fundamental pillars, all drawing from important breakthroughs in the broader AI community.

- Pillar 1: Deep Learning Fundamentals. The frequent co-citation of foundational works like Goodfellow et al. (2014) on Generative Adversarial Networks (GANs) demonstrates that PCGML is a direct application of modern deep learning theory. The centrality of these papers confirms their status as the bedrock of the field.
- Pillar 2: Deep Reinforcement Learning. The second pillar is defined by the influential works on deep reinforcement learning, such as Mnih et al. on Deep Q-Networks and Schulman et al. on Proximal Policy Optimization (PPO). Their prominence signifies that creating agents capable of learning through interaction is a central methodology in PCGML.
- Pillar 3: The third pillar is unique, represented by the highly-cited survey by Summerville et al. (2018). While not a foundational AI paper itself, it

plays a crucial role as an intellectual bridge. It translated the powerful theories from the first two pillars into the specific context and language of game development, thereby legitimizing PCGML as a distinct field of study.

In summary, while the social structure of PCGML is defined by its main cluster, and those at the borders, collaboration network, its intellectual structure is firmly linked to deep learning and reinforcement learning breakthroughs of the 2010s.

CHAPTER 5
DISCUSSION & CONCLUSION

This final chapter serves to synthesize the bibliometric findings presented in Chapter 4, while contextualizing them within the broader landscape of artificial intelligence and the video game industry. The objective is to present the key insights from this study along with their practical and theoretical implications for both academics and industry practitioners, while acknowledging the limitations of the research, and propose a clear path forward for future inquiry. In essence, this chapter connects the data-driven analysis with the real-world problem statement that motivated this thesis: the need for a clear, evidence-based map of the fragmented but rapidly evolving AI-driven Procedural Content Generation (PCGML) research field.

5.1 Summary of the Key Findings

This study was conducted to map the intellectual, social, and conceptual structure of PCGML research from 2008 to 2025. The analysis of 2,184 curated documents has yielded several insights that provide a deeper understanding of the field.

- A Field Defined by Three Time Periods: The evolution of PCGML (along with AI development) is not linear but is marked by three distinct periods. It began with the "Age of Classic AI" (2008-2015), a time focused on search algorithms and agent-based problem-solving. This was followed by the "Machine Learning Revolution" (2016-2021), where data-driven methods like reinforcement learning and neural networks became the dominant "motor themes." The current period of generative AI (2022-Present) is characterized by large-scale generative models capable of sophisticated content creation.
- A Symbiotic Ecosystem of Academia and Industry: The primary centres of research are not confined to universities. While academic powerhouses like New York University lead, they operate alongside and in collaboration with major corporate research labs, including Tencent

AI Lab, Netease Fuxi AI Lab, and Microsoft Research. This demonstrates that PCGML is not a purely theoretical pursuit but a domain with commercial interest.

- A Disproportionate Research Impact: While the United States and China are the undisputed leaders in the sheer volume of research, the United Kingdom demonstrates a disproportionately high scholarly impact, with an average of 13.64 citations per document. This is nearly an order of magnitude higher than most other leading nations and suggests a research strategy potentially focused on foundational or high-risk, high-reward studies that become cornerstone references for the entire field.
- A "Core-Periphery" Social Structure: The field's productivity is heavily influenced by a "core-periphery" social structure. A highly interconnected central research group, centered around influential authors like Julian Togelius and his collaborators, forms the field's intellectual "center of gravity." This core is complemented by numerous smaller, independent research groups that ensure a diverse and healthy ecosystem with more experimental work done on the outskirts of the field while the main centres work on conventionl topics.
- Anchored in Foundational AI Breakthroughs: The intellectual bedrock of PCGML is firmly anchored in the important breakthroughs of the broader AI community. Co-citation analysis reveals that the entire field stands on three pillars: the Deep Learning Fundamentals (e.g., Goodfellow's work on GANs), the Deep Reinforcement Learning revolution (e.g., Mnih's work on Deep Q-Networks), and a crucial Paper (Summerville et al., 2018) that translated these powerful AI concepts for a game development audience.

5.2 Implications and Recommendations

The findings of this study offer actionable insights for two key groups: the academic community and industry practitioners.

5.2.1 Academics: Implications and Recommendations of Study

Roadmap for New Researchers: This analysis serves as a guide for new scholars, helping them quickly identify the fundamental papers, key methodologies, and most active research hubs. It provides a clear picture of the state-of-the-art research topics being done, reducing the risk of redundant research.

Identifying Research Gaps: The thematic map highlights emerging themes in the bottom-left quadrant (e.g., extended reality, computer vision). These represent early research areas with lower density, signaling opportunities for original contributions.

Fostering Collaboration: The visualization of collaboration networks and the clear presence of industry labs can help researchers identify potential partners outside their immediate circle, bridging the gap between the core, the periphery, and industry to foster a more integrated research community.

5.2.2 Business/Industry: Implications and Recommendations of Study

Strategic Technology Adoption: The timeline of technological trends provides a data-driven guide for R&D investment. The clear progression from classic AI to deep generative models validates strategies focused on modern, data-driven content creation tools.

Talent and Collaboration Scouting: The ranked lists of the most productive authors, institutions, and corporate labs serve as a directory for recruitment and partnership. Companies seeking to build their AI teams can identify the universities and research groups at the forefront of innovation.

Staying Ahead of the Curve: The analysis of publication venues confirms that to remain current, one must monitor top-tier, fast-moving conference proceedings from venues like IEEE, ACM, and AAAI, rather than relying solely on the slower timeline of traditional journals.

5.3 Limitations of the Research

While this study provides a robust overview, it is essential to acknowledge its limitations to ensure a balanced interpretation of the findings.

Database Scope: The analysis was exclusively confined to the Scopus database. Seminal works published in other sources or non-peer-reviewed but influential sources (e.g., arXiv, industry blogs) were not included.

Keyword Dependency: Bibliometric analysis relies on the keywords, titles, and abstracts provided by authors. It cannot capture the full degree or implied meaning within the body of a paper, which may lead to underrepresentation of certain concepts.

Quantitative Focus: This study is, by design, quantitative. It maps what is being researched and by whom, but it cannot make qualitative judgments about the quality or correctness of individual papers.

Language Bias: The exclusion of non-English publications means this study reflects the English-language research landscape. Significant contributions from other linguistic communities may have been missed.

5.4 Future Research and Next Steps

This study lays the groundwork for several promising avenues of future research that can build upon its findings and address its limitations.

- Qualitative Research: A natural next step is a qualitative systematic review focusing on one of the identified thematic clusters, such as the "emerging themes" of extended reality, to complement the scope of this study with analytics.
- Cross-Database Comparison: A future bibliometric study could compare findings from Scopus with those from Web of Science or Google Scholar to create an even more comprehensive map of the field.
- Impact on Game Production Practice: A study employing methodologies like industry surveys or case studies could investigate the real-world adoption rates of the academic trends identified here, uncovering the practical barriers to implementation.

5.5 Concluding Remarks

The field of AI-driven Procedural Content Generation is at a critical moment. It has evolved from a niche academic interest into a powerful engine of innovation, fundamentally reshaping the creative and economic landscape of the video game industry. This bibliometric analysis has sought to replace fragmented knowledge with a clear, data-driven map. The findings reveal an exciting, rapidly maturing field characterized by a symbiotic relationship between academia and industry, a clear technological journey towards sophisticated generative models, and a social structure that balances focused, core research with broad, exploratory inquiry. By understanding the key contributors, dominant themes, and collaborative networks that define its past and present, both researchers and practitioners are better equipped to navigate its future, fostering the next generation of dynamic, personalized, and endlessly replayable interactive experiences.

BIBLIOGRAPHY

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R tool for comprehensive science mapping analysis. *Journal of Informetrics,* 11(4), 959–975.

El-Nasr, M. S., Drachen, A., & Canossa, A. (Eds.). (2021). Games, Learning, and Society: The Global and the Local. *ETC Press.*

Hendriks, K. (2023). Procedural Content Generation for Games: A practical introduction to creating worlds. *CRC Press.*

Jain, A., Isaksen, A., & Togelius, J. (2020). DRL-RPG: A reinforcement learning-based framework for role-playing games. *2020 IEEE Conference on Games (CoG),* 461–468.

Millington, I., & Funge, J. (2009). Artificial intelligence for games. *CRC Press.*

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* 372.

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

Short, T., & Adams, T. (2017). Procedural content generation in games. *CRC Press.*

Smith, G. (2015). The future of procedural content generation. *Proceedings of the 10th International Conference on the Foundations of Digital Games.*

Statista. (2023). Video Games - Worldwide. Retrieved from https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide

Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A. K., Isaksen, A., ... & Togelius, J. (2018). Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games,* 10(3), 257–270.

Volz, V., Schrum, J., Liu, J., Lucas, S. M., Smith, A., & Risi, S. (2018). Evolving Mario levels in the latent space of a deep convolutional generative adversarial network. *Proceedings of the Genetic and Evolutionary Computation Conference,* 221–228.

Yannakakis, G. N., & Liapis, A. (2017). Experience-driven procedural content generation. *IEEE Transactions on Affective Computing,* 8(3), 289-303.

Yannakakis, G. N., & Togelius, J. (2018). Artificial intelligence and games. *Springer.*

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. *Springer.*

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development,* 3(3), 210-229.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, 83(1), 243-258.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics,* 5(1), 148-166.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research,* 133, 285–296.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), Advances in neural information processing systems 27 (pp. 2672–2680*). Curran Associates, Inc.*

Pritchard, A. (1969). Statistical bibliography or bibliometrics?. *Journal of Documentation,* 25(4), 348–349.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. *The MIT Press.*

BIODATA

Name-Last Name:	Ralf William Alexander Schmidt
Email:	Williamschmidt91@gmail.com
Education Background:	Bachelor of Economics (Xiamen University)
Work Experience:	Five years of experience working in a 3D art outsourcing company as Operations Manager in the videogame industry, working with companies such as: 4A Games, Ubisoft, Saber Interactive and Paradox Entertainment.