https://doi.org/10.51408/issi2025_163

Citation Context Analysis: Evaluating Human vs. AI Annotations in Gameplay Bricks Research

Marc Bertin[1], Julian Alvarez[2], Thierry Lafouge[3]

[1]*marc.bertin@univ-lyon1.fr*
ELICO, Université Lyon1, Lyon, Villeurbanne (France)

[2]*julian.alvarez@univ-Lille.fr*
GERiiCO, Université de Lille, Lille (France)
Immersive Factory, R&D, Paris (France)

[3]*thierry.lafouge@univ-lyon1.fr*
ELICO, Université Lyon1, Lyon, Villeurbanne (France)

Abstract

Recent advances in citation analysis have moved beyond traditional bibliometric approaches to explore the contextual roles of citations in academic discourse. While Large Language Models (LLMs) offer new possibilities for analyzing citation contexts, challenges persist regarding annotated dataset availability and inherent biases in citation categorization schemes. This study presents a novel comparative analysis of citation contexts, focusing on the gameplay bricks framework developed by Alvarez and Djaouti (2006) across a ten-year period (2008-2018).

Our research employs prompt engineering techniques to analyze citation contexts in nine languages, comparing human expert annotations with ChatGPT-generated analyses. This micro-level investigation examines how the gameplay bricks model has been referenced, appropriated, and critiqued across different disciplines. The study addresses three primary research questions: the interpretation of citation contexts by domain experts, the alignment between AI-generated categorizations and expert judgments, and the insights gained from comparing human and AI annotations in multilingual scientific discourse.

The methodology combines traditional human annotation with AI-assisted classification through prompt-based methods. Our analysis reveals a predominance of definition and appropriation categories, indicating widespread adoption of the gameplay bricks model across disciplines. Computer science publications showed higher rates of model appropriation, while humanities disciplines demonstrated more critical engagement. The study identified particular challenges in capturing neutrality and criticism, attributable to both AI model limitations and the inherent complexity of citation context interpretation.

Results demonstrate that while ChatGPT-powered annotation offers scalability advantages, it faces limitations in processing contextual nuances and interpretive depth, particularly across different languages. The comparative analysis highlights discrepancies between human and AI interpretations, suggesting the need for hybrid approaches that leverage both human expertise and AI capabilities. These findings contribute to ongoing discussions about AI's role in academic discourse analysis and raise important questions about citation practices.

It provides insights into the evolution of academic discourse around the gameplay bricks framework while highlighting methodological considerations for future citation analysis studies. The findings underscore the importance of developing more sophisticated tools for citation context analysis that can account for linguistic and disciplinary variations.

Introduction

Since a comprehensive theory of citation has not yet been established, emerging models are being proposed, as highlighted by Tahamtan & Bornmann (2022) with *The Social Systems Citation Theory (SSCT): A proposal to use the social systems theory for conceptualizing publications and their citation links*. Research on citation contexts focuses on two interrelated and complementary aspects: a conceptual dimension that leads to the proposal of schemas, categories, and functions related to the nature of citation acts, and an operational dimension of these categories, which requires the implementation of corpora, computational modeling, and annotation evaluation.

Today, numerous citation categories have been proposed and many surveys written such as Bornmann & Daniel 2008, Hernández-Alvarez & Gomez 2016, Zhang et al 2023 to name but a few, and computational tools are becoming more powerful with the advent of large language models (LLMs). This paradigm shifts in methods derived from natural language processing (NLP) opens new perspectives. However, two main limitations remain: the scarcity of annotated resources for training machine learning methods and the nature of the categories on which supervised approaches rely. For this article, drawing on recent advancements in prompt engineering, we propose a case study to explore the relevance of analyzing citation contexts over a ten-year period for a specific research topic: the *gameplay bricks* framework introduced by Julian Alvarez and Damien Djaouti in 2006. Our study focuses exclusively on the analysis of citation contexts from a temporal perspective. Understanding citation contexts requires examining their evolution over time. At a micro-level—opposed to the macro and meso approaches of traditional bibliometric studies—returning to the text and conducting a fine-grained analysis are essential for understanding controversies and debates. For example, a dispute between two researchers through successive articles can only be analyzed by reading the texts in full, even if the citation frequency is low. A highly cited article, on the other hand, indicates high visibility, meaning it appears in many bibliographies. But what role does it play for the citing researcher? Why and where was it cited? Studies have shown that the rhetorical structure of a scientific article significantly influences the nature of citation contexts, depending on whether the citation appears in the introduction (literature review), methodology, results, or discussion sections.

The study we propose confronts human expertise—represented by an identified researcher in the field of gameplay studies—with the latest AI approaches using prompts, specifically ChatGPT, to analyze citation contexts within this corpus. To illustrate our approach, we selected a case study based on the *gameplay bricks* framework introduced by Julian Alvarez and Damien Djaouti in 2006. The advantage of this work is that it includes an inventory of international citations, which have been analyzed by one of the researchers to identify citation contexts. The goal is to determine whether these citations reflect an adoption of the *gameplay bricks* model, a critique of it, or a neutral stance (Alvarez, 2018). This provides a basis for a comparative study between human expertise and AI analysis. Additionally, this corpus offers other advantages, such as its multilingual nature, making it suitable for a first iteration of a comparative study between human and AI analyses.

Our approach thus leads us to examine the nature of citations received by various research articles. For example, we may explore the fundamental nature of citations related to gameplay studies: Are they negative, positive, or neutral? Can we identify cases of conceptual appropriation, and if so, of what kind? Which scientific fields refer to the studied works? Are these works cited in other languages? In the latter case, what functions do the citation contexts convey?
We have therefore chosen to compare the analysis performed by a researcher with that of an AI on the same corpus to better understand the identified and generated citation contexts in both cases. Beyond the question of reliability, we also consider it relevant to leverage such a comparative analysis to uncover the insights that such a cross-analysis can provide. This constitutes the primary objective of this article. In the context of using AI to help identify citation contexts, the underlying issue of reliability will be addressed in the evaluation section.

GPT and LLM Litterature review

Over the past few decades, automatic classification of citation features has evolved in parallel with advances in natural language processing (NLP) technologies. However, despite numerous studies documented in surveys (Bornmann and Daniel (2008); Hernández-Alvarez and Gomez (2016); Jha, Jbara, Qazvinian, and Radev (2017); Lyu, Ruan, Xie, and Cheng (2021), significant limitations and persistent biases hinder its widespread adoption. One of the most noteworthy advancements in natural language processing (NLP) is the emergence of large language models, such as GPT (Generative Pre-trained Transformer) (Radford (2018), 2019; Brown et al. (2020a)), along with their numerous iterations. These models have exhibited exceptional capabilities across a diverse range of linguistic tasks. Typically, GPT models undergo two key phases: pre-training on extensive text datasets to learn general language patterns, followed by fine-tuning for specific downstream tasks to generate highly human-like language. Among the various fine-tuning techniques, prompt engineering stands out as a particularly accessible approach for non-specialists, offering a user-friendly means to harness the potential of these powerful models. While Nishikawa's research demonstrated the consistency of LLMs in this task, it also highlighted the limits of their ability to fully replace human annotators. Indeed, Lahiri et al. (2023) introduce CitePrompt, a novel tool leveraging prompt learning for citation intent classification. By optimizing the choice of pretrained language models, prompt templates, and verbalizers, CitePrompt achieves state-of-theart performance on the ACL-ARC dataset and significant improvements on SciCite, requiring minimal external document information. They propose a first-of-its-kind approach to adapt citation intent classification to few-shot and zero-shot settings, addressing the scarcity of large labeled datasets.

Zero-based and low-based learning for labeling citation contexts

Nevertheless, emerging approaches such as Zero-Shot and Few-Shot Learning for citation labeling, inspired by the work of Brown et al. (2020b), offer promising avenues for exploration. In fact, the literature shows that other fields use this type of approach to compensate for the lack of annotated corpora. ChatGPT offers a broad

spectrum of applications in the field of research, particularly in the domain of text mining. For example, Mathebula, Modupe, and Marivate (2024) in sentiment analysis for financial applications, enhancing the accuracy and utility of customer feedback in shaping business decisions. Khan, Khan, Li, Ullah, and Zhao (2025) introduces a novel approach using ChatGPT as both annotator and negotiator, achieving a 94% accuracy rate with deep learning classifiers in detecting emotions in negative reviews from low-rated apps, demonstrating the potential of generative AI in enhancing annotation reliability and performance. Chen et al. (2023) evaluates ChatGPT's performance on biomedical tasks through a comprehensive benchmark involving article abstracts, clinical trial descriptions, and biomedical questions, demonstrating its effectiveness and versatility in biomedical text comprehension, reasoning, and generation. Zhu et al. (2022) have described the fundamental concepts underlying this approach, which could play a central role in advancing citation analysis. More recently, Lahiri, Sanyal, and Mukherjee (2023) have positioned Prompt Learning as a particularly suitable method for tackling this challenge.

Discussion of Gameplay Bricks Model

The distinction between the concepts of video games and Serious Games is based on principles modeled by Alvarez, Djaouti, Ghassempouri, Jessel, and Methel (2006). This model, named "Gameplay Bricks", was originally designed to deconstruct video games in an effort to both classify video games and identify characteristics that could distinguish Serious Games from video games within a formal system Alvarez et al. (2006). After 2006, the Game Bricks model was consolidated over the period from 2007 to 2010. The literature on which this model is based is presented in Table 2. While the core of the model will be repeated in the literature, it is interesting to note the variety of media used to build the Game Bricks model. More than a decade after its introduction into the scientific community, how has the Gameplay Bricks model been perceived, used or criticized? What specific criticisms can we identify from the citations collected? The corpus for this study will come to an end in 2018, when a synthesis book will be published on this issue. In 2024, we will have the necessary hindsight and coverage to observe the spread of this model within the various scientific communities. Indeed, the choice of this model is even more interesting in that it is mobilized through numerous national and international citations, in different languages and in different contexts. The game bricks expert was able to build up a categorization of citation functions through manual study and human expertise.

Problems

Furthermore, the underlying question that interests us in this study is whether, given the current state of research on citation contexts, we are capable of producing semantic annotations of citation contexts that would ultimately allow us to track the dissemination of models, as demonstrated in this study, or theories, as well as the identification of controversies. Our research problem is as follows: Based on the corpora generated around the modeling of Game Bricks, can we analyze citation contexts and derive categories that align with expert-produced knowledge?

As highlighted in the literature review, we face two major limitations: the lack of stable categorizations and a bias introduced by supervised approaches, which still lack annotated corpora covering all categories and disciplines. Recent state-of-the-art reports shed light on the latest studies in this field. The Gameplay Bricks model is employed to determine whether ChatGPT-based approaches applied to citation contexts can provide an application framework for understanding discussions, or even controversies, surrounding this model. To assess ChatGPT's potential in research and its application to citation contexts, our study explores its understanding of semantic usage, focusing on specialized topics related to gameplay bricks. Based on these elements, we propose the following three main research questions:

1. How does a domain expert mobilize citation contexts?
2. Do the categorizations produced by ChatGPT agree with the expert?
3. How to navigate in a contextualized space of citation contexts?

The content of this paper is organized as follows: Section 1 provides a general introduction, including a state-of-the-art review and the research problem addressed. Section 2 describes the dataset constructed around Gameplay Bricks and outlines our experimental approach to citation context categorization, which integrates human expertise with the proposed solution using OpenAI ChatGPT. Section 3 presents the results obtained from AI-generated citation context annotations, the resulting graph, and a human analysis focusing on cases of appropriation. Section 4 offers a discussion comparing the human analysis with the proposed OpenAI ChatGPT approach.

Methodologies

As we have just observed, the potential of this type of approach is evident. Two studies that have particularly drawn our attention in constructing our methodology are the studies of Lahiri, A., Sanyal, D.K., Mukherjee, I. (2023) and the latest research from Nishikawa, K., & Koshiba, H. (2024), which explores the application of large language models (LLMs) to citation context analysis. The article of Nishikawa, K., & Koshiba, H. (2024) highlights a crucial limitation of current approaches: the lack of annotated corpora. The study emphasizes the experiment's inability to achieve relevant annotation results. This research employed five classes for *Citation Purpose* when citing a referenced paper: Background, Comparison, Critique, Evidence, and Use. Regarding *Citation Sentiment*—which refers to the mental attitude of the author of a citing paper towards the cited paper—the authors used three classes: Positive, Negative, and Neutral. The choice of categories is based on the availability of an annotated corpus and resources for the scientific community.

Designing Prompts for Citation Context Classification

One of the key aspects of this approach with Large Language Models (LLMs) is the design and application of prompts, which are structured natural language inputs that guide the model's response. In this context, the structure and specificity of the prompt significantly influence the quality and relevance of the generated output. To ensure

optimal performance, it is essential to provide input categories that do not involve intrinsic complexity or rely on excessively broad generalizations. The prompt must explicitly instruct the LLM to focus on classification, ensuring that the model produces the expected results. To further enhance prompt precision, several strategies can be employed. One effective approach is to provide explicit examples of citation contexts. Instead of relying on a single query, using labeled citation contexts allows the model to generalize more effectively and improve the relevance of its outputs. Still in the context of the implementation of our method, we have taken on board the remarks of Nishikawa 2024 concerning techniques for improving results, namely the few-shot Brown et al. (2020) or chain-of-thought approaches of Wei et al. (2022) and Zhang (2023) fot Chain-of-thought with ChatGPT for Stance Detection on Social Media.

Nishiwaka's methodological approach is structured around four citation incentive models that build upon the basic instructions by progressively incorporating additional contextual elements. The first model includes only class types, providing a minimal framework for classification. The second model expands on this by integrating class types along with their definitions, offering greater conceptual clarity. The third model further refines the approach by including annotation procedures, ensuring more precise and standardized applications. The most comprehensive model incorporates class types, definitions, annotation procedures, keywords, and example sentences, creating a fully detailed framework for citation analysis. This final model closely resembles the manual used in their previous study (Nishikawa, 2023).

In line with these results, it is essential to develop a prompt-based approach that takes these constraints into account. To achieve this, we need to produce a classification for citation contexts that provides a clear and structured framework for annotation. This leads to the next point, which is to identify the categories that can meet the prompting requirement.

Construction Citation Context Classification

Previous studies provide a certain richness despite inherent biases in their construction and design, such as corpus size, disciplinary scope, or unaddressed biases, such as the language used. We can cite the work of Teufel (2006), Athar (2011), Dong & Schäfer (2011), Bertin & Atanassova (2024) as key references for this study. Our approach focuses on the various categories proposed in the literature to design prompts that provide classifications aligned with the discursive forms likely to appear in citation contexts. These classifications aim to minimize ambiguity and abstraction while accurately reflecting the nature of citations. To achieve this, we draw on the work of Liu et al. (2023), and its applications in identifying citation intents in scientific papers, as explored in the recent study (Nishikawa, K., & Koshiba, H., 2024).

The categories used to build the chatGPT prompt naturally draw on the labels proposed by the expert, but also on other categories identified in the literature. The categories selected must convey notions that can be identified in a citation context. To this end, we have selected categories identifiable by discourse forms likely to be

present in citation contexts. Based on established classifications, we have produced a selective and descriptive list of functions. These functions are likely to be relevant to the processed corpus and implementable via prompts for the intended task. We have synthesized the main citation functions, as outlined in the following table 1:

Table 1. A Synthesis of Citation Functions: Categories and Their Discursive Roles.

Category	Description
Definition	The citing paper provides a definition of a concept from the cited paper.
Method	A citation instance where the cited work provides the method or technique used in the citing paper, which either describes or applies the methodology introduced in the cited work.
Hypothesis	The cited work is used to support or inspire a hypothesis in the citing work.
Extension	The citing work's research work is an improvement or extension of the cited work.
Comparison	A citation instance that involves any form of comparison or contrast between different cited papers or between the cited work and the citing paper. It highlights similarities or differences between the cited work and the author's own research.
Agreement	The citing paper explicitly agrees with or endorses the cited paper's conclusions.
Result	A citation instance in which the citing work mentions specific results or general findings of the cited paper.
Extension	The citing paper extends the methods, tools, or data of the cited paper.
Point of view	The cited work is used to illustrate a particular theoretical or conceptual perspective
Future	The cited paper may be a potential reference for future work.

The Gameplay Bricks Corpus

The distinction between video games and Serious Games is based on principles formalized by Alvarez et al. (2006) through the "Gameplay Bricks" model. Initially developed to deconstruct video games, this model aimed to establish a classification system while identifying specific characteristics that differentiate Serious Games from traditional video games (see Alvarez et al., 2006). Following its introduction, the Gameplay Bricks model was further refined between 2007 and 2010. Table 2 provides a synthesis of the literature that contributed to the development of this model. While its foundational principles are consistently referenced in subsequent research, it is particularly noteworthy that a diverse range of media has been utilized in shaping and expanding the Gameplay Bricks framework.

Table 2. Works published between 2006-2010.

Categories	References
Conference	Alvarez J., D. Djaouti, and R. Ghassempouri (2006), "*Morphological study of videogames*," CGIE'06 conference, Australia.", 2006
Conference Proceedings	Djaouti, Damien, J Alvarez, Jp Jessel, Gilles Methel, and P Molinier. 2007. "*The Nature of Gameplay: A Videogame Classification*." Cybergames Conference, no. July 2015.", 2007
Conference Proceedings	Djaouti, D., Alvarez, J., Jessel, J.-P., & Methel, G. (2007). *Towards a classification of video games*. Artificial and Ambient Intelligence convention (Artificial Societies for Ambient Intelligence) (AISB (ASAMi) 2007).", 2007
Conference Proceedings	Alvarez et al., 2007] Alvarez, J., Djaouti, D., Jessel, J.-P., Methel, G. et Molinier, P. (2007). *Morphologie des jeux vidéo*. In H2PTM, Hammamet, Tunisie, 29/10/2007-31/10/2007, numéro 978-2-7462-1891-8 de Lavoisier, pages 277–287, http://www.editions-hermes.fr/. Hermès Science Publications.", 2007
Thesis	Thesis,"Alvarez, J. (2007). *Du jeu vidéo au serious game, approches culturelle, pragmatique et formelle*, Thèse de doctorat en science de l'information et de la communication, Toulouse, France : Université de Toulouse.", 2007
Article	Djaouti, D., Alvarez, J., Jessel, J.-P., and Methel, G. (2008). *Play, Game, World: Anatomy of a Video-Game*. International Journal of Intelligent Games & Simulation, 5(1):35–36.", 2008
Book	Book,"Alvarez, J., & Djaouti, D. (2010), "Introduction au Serious Game", Questions théoriques, vol. 1, Paris.", 2010
Website	Website,"Alvarez, Julian et Damien Djaouti. S.d. Game Classification: la classification en ligne du jeu vidéo. <http://www.gameclassification.com/>.", 2010

The first step was to create an average from the WoS based on the search equation to build a corpus: Gameplay Bricks (All Fields) and CMN-3138-2022 (Auhtor Identifiers) or AAE-9793-2019 (Author identifiers) which produced 6 references for 45 citations from 2007 to 2023. For this study, which covers the period from 2008 to 2018, the Web of Science (WoS) database reports 3 articles with a total of 19 citations. From this equation, the results were extended via other databases to cover the multilingual aspect. As the concept is mobilized by the international community and has a coverage that goes beyond English-language publications, it was important to extend our research to have a consolidated corpus. We identified a total of 47 scientific articles in 9 languages and 40 theses in 4 languages, highlighting the richness and international scope of the concept explored in this study, using additional resources, databases, as well as laboratory and institutional websites. Using Google Scholar with keywords such as Brique Gameplay and Gameplay Brick combined with the names Djaouti or Alvarez, nearly 200 national and international references were identified in 2018. Among these results, self-citations were removed, ensuring that the same author was cited only once, with preference given

to the oldest or most detailed article referring to the Gameplay Brick model. In addition, articles mentioning the notion of bricks without referring to or using the model were excluded. Indeed, the term brick is often used in everyday language to refer to the idea of a component. This process resulted in a final count of 47 articles explicitly citing the Gameplay bricks model. Regarding Ph.D. theses, we have identified 16 in English, 2 in Spanish, 20 (including HDRs) in French, and 2 in Portuguese. We conducted a detailed analysis of the metadata of the corpus, which we present below in the various tables.

Table 3. Distribution of the Number of Articles by Year.

Years	Number of Articles	Percentage
2008	1	2.1 %
2009	3	6.4 %
2010	3	6.4 %
2011	6	12.8 %
2012	4	8.5 %
2013	3	6.4 %
2014	3	6.4 %
2015	9	19.1 %
2016	11	23.4 %
2017	3	6.4 %
2018	1	2.1 %
Total	**47**	**100%**

Table 4. Distribution of the Number of Articles by Discipline.

Discipline	Number of Articles	Percentage
Computer Science	23	48.9%
Education	8	17.0%
Art	6	12.8%
Information Sciences	5	10.6%
Industrial Engineering	1	2.1%
Management	1	2.1%
Language / Literature	1	2.1%
Philosophy	1	2.1%
Health	1	2.1%
Total	**47**	**100%**

Table 5. Distribution of Authors by Nationality.

Nationality	Number of Authors
Germany	6
Belgium	3
Brazil	1
Bulgaria	5
Canada	4
Korea	2

Denmark	1
Spain	2
Estonia	1
France	19
Greece	1
Ireland	5
Italy	1
Japan	4
Netherlands	1
Portugal	1
Morocco	12
Mexico	5
United Kingdom	9
Russia	3
Sweden	3
Switzerland	1
Taiwan	4
USA	3

Table 3 illustrates the temporal coverage of our corpus, spanning the period from 2008 to 2018. Table 4 provides information about the disciplines identified based on the journals or conferences in which the scientific articles were published. Another aspect we considered relevant was the identification of authors and their nationalities. The data obtained is presented in Table 5.

The Gameplay Bricks Full Text Dataset

The corpus is primarily composed of PDF documents. These were converted into text for an initial preprocessing phase, enabling the analysis of the language used in scientific articles, PhD theses, and habilitation theses that were identified during our bibliographic research. The corpus used in this study consists exclusively of scientific articles, based on the full-text content extracted from PDF documents. The analysis of PhD theses will be addressed in future research and will be discussed in the context of the creation of new knowledge.

GROBID is a machine learning library designed to extract, parse, and restructure raw documents, such as PDFs, into structured XML/TEI documents. It is particularly suited for processing technical and scientific publications. In our study, we utilized the GROBID Web API, which provides a straightforward and efficient interface to the tool. The service was deployed within a Docker container running on Linux. For processing documents, we used the associated Python client, enabling concurrent processing of a batch of PDF files located in a specified directory. The experiments were conducted on a machine featuring an Intel® Core™ i7-4790K (8 threads) processor and 32 GB of RAM. No specific optimizations were applied to the GROBID processing pipeline, as the corpus size did not warrant such measures. GROBID was configured to generate TEI files with options tailored to the needs of our study. Specifically, the tool was set to perform sentence segmentation in the TEI XML output. This segmentation leverages the OpenNLP sentence detector, which is

recommended for scientific articles. The TEI generated by GROBID establishes a link between citation contexts and bibliographic references, enabling the construction of a matrix of relationships between citing and cited references. This approach allows us to connect the semantic categories produced by humans and machines in a network. The network will be a directed graph, with a label corresponding to the semantic category. For that purpose, we used Gephi to propose a practical case of visualization of the semantic network of games play from the labelled corpus Bastian, Heymann, and Jacomy (2009).

The following Table 6 shows a summary of the data processing produced by GROBID and corrected to produce a multilingual dataset of citation contexts to be explored. Indeed, the multilingual aspect poses difficulties in the conversion to TEI. We had to make corrections to improve context coverage. Nevertheless, the corrections we have made enable us to build a dataset referencing the founding articles of games bricks, and consequently to propose the dataset desired by our approach.

Table 6. Distribution of Articles, Citation Contexts, and References by Language.

Languages	Number of Articles	Number of Processed Articles	Number of Citation Contexts	Number of References in the Reference Corpus
English	44	44	962	58
Korean	2	2	49	3
Spanish	1	1	30	1
French	20	20	540	28
Indonesian	2	2	78	n.d.
Persian	1	n.d.	n.d.	n.d.
Portuguese	1	1	26	2
Russian	3	3	165	4
Swedish	1	1	68	1
Thai	1	n.d.	n.d.	n.d.
Total	**76**	**74**	**1918**	**97**

 n.d. not determined

Gameplay Bricks Labeling: A Human-Centric Perspective

The human approach was conducted in April 2018, relying in particular on Google Scholar via the use of the keywords “Brique Gameplay” and “Gameplay Brick” by associating the names “Djaouti” or “Alvarez” (Alvarez, 2018: pp42-73). A recent search carried out in 2023, again using Google Scholar, revealed 33 additional references for the same period. The corpus studied with the human approach thus represents 79 documents. The documents are then classified according to the type of

citation: The expert defined three labels in order to respond to his problem without taking into account existing categories: "Neutral", "Critic" and "Appropriation" (cf. Table). The Critic and Appropriation criteria are not mutually exclusive. Indeed, appropriation does not necessarily mean that the author expresses no criticism of the model. Some authors, like Pierre-Yves Hurel, take the trouble to criticize the model in order to appropriate it later on: To establish our own typologies (types of actions, types of rules), we propose to present and criticize the theory of gameplay bricks. As we shall see, this concept, which was created with the aim of improving game classification, can give us the tools we need for ideological analysis (Hurel, 2011, p29). With this in mind, it is worth drawing criticism also from the writings of researchers who have appropriated the model.

Cases of Gameplay Brick appropriation

In this subsection, the idea is to present the different types of appropriation of Gameplay Bricks identified by researchers and presented in Table 7. A dozen articles present an appropriation among which four types of appropriation can be identified. Five types of appropriation of the Gameplay Bricks model were identified in 2018. We'll take a closer look at these different types in the following subsections.

Table 7. Categorization of labels in the context of game bricks by a human expert, based on their purpose and knowledge of the field.

Labels	*Description of labels in the context of games bricks*	
Neutral	means that the Gameplay Bricks are merely cited by the article, but the author expresses no opinion on the model	
Critic	indicates that the article will significantly point out limitations or a disagreement with the model	
Appropriation	denotes a consideration of the model in the author's work.	
	Type 1 (T1)	**Use model:** Appropriation concerns the use of Gameplay Bricks to design or deconstruct Serious Games or video games.
	Type 2 (T2)	**Inspire methodologies:** Identified appropriation draws inspiration from the Gameplay Bricks to build new methodologies.
	Type 3 (T3)	**Integrate model:** The appropriation identified represents the integration of the Gameplay Bricks model into other models.
	Type 4 (T4)	**Develop experiments:** The appropriation of Gameplay Bricks is linked to the development of scientific experiments. However, in a more recent analysis of the additional references identified since 2023, we have identified a 5th type
	Type 5 (T5)	**Justifying a theoretical approach:** The appropriation of Gameplay Bricks is linked to a theoretical construction.

Appropriations of type 1: Use model

The first type of appropriation identified in the research literature concerns the use of Gameplay Bricks to deconstruct existing Serious Games or video games, or to help the design of new ones. This is the intended use when the model was developed. In this respect, we refer in particular to the article by Carlos Delgado-Mata, Ricardo Ruvalcaba-Manzano, Oscar Quezada-Patino, Daniel Gomez-Pimentel and Jesus Ibanez-Martinez: For the video game developed for this research, the bricks of interest are Move, Avoid and Reach. These types of bricks are well suited to our objective of developing a game that measures and develops fine and gross motor skills (Delgado-Mata, Ruvalcaba-Manzano, Quezada-Patino, Gomez-Pimentel, & Ibanez-Martinez, 2009, p5).

Type 2 appropriations: Inspiring methodologies

The second type of appropriation identified encompasses work that draws inspiration from the Gameplay Bricks to build methodologies. Marion Coville explains how she appropriated the Gameplay Bricks to build her experimental methodology for studying issues of gender, representation and role in video games (Coville, 2011, p 165). The researcher explains:

My methodology is based on this classification. First of all, I list the rules and actions available in the games, as well as the objectives and relationships to the world and universe in which the character evolves. I do this through my own experience of the game, while paying particular attention to the testimonies of other players. Once the modalities of interaction between the game and the player have been identified, I turn to the representation of heroines (Coville, 2011, p 172).

Type 3 appropriations: Integrating models

The third type of appropriation identified represents the integration of the Gameplay Bricks model into other models. This is the case, for example, of Yuri Gomes Cardenas, who proposes an ontology model designed to represent Serious Video Games. Among the elements that make up his model, the Gameplay Bricks model is thus mobilized (see Cardenas et al., 2014, p85)

Type 4 appropriations: Designing experiments

The fourth type of appropriation is linked to the development of scientific experiments. This is the case of Gaël Gilson, who proposed an experiment to study whether a gamer's virtual experience could represent an informal learning situation. One of the aims of the protocol was to ask subjects to identify the Gameplay Bricks they thought they would mobilize during the video-game activity, in order to understand how they ultimately they fit into the activity and the links they might establish with potential learning. The part of the protocol that calls upon the Gameplay Bricks is initially explained in the form of texts that are comprehensible to young subjects (Gilson, Draelants, Jardon, & Servais, 2016, p186). Once the subjects have been interviewed, the data collected is mapped (Gilson et al., 2016,

p187). Gameplay bricks are then listed in the same way as in the original English-language model, in the column Gameplay bricks employed.

Type 5 appropriations: Justifying a theoretical approach

This fifth type of appropriation aims to mobilize the Gameplay Bricks model to conduct a theoretical demonstration or corroborate theoretical approaches. This proposal does not intend to classify games, but to catalogue elements within a hierarchical structure. This catalogue can be used to describe the game according to its design space. It can also work as a framework to explore research questions related to games and gameplay, as proposed by the gamebricks classification, or to construct a vocabulary for describing, analyzing and critiquing games.

Results

Discipline and positioning overview

Based on the data listed in Table 2, Table 4 shows, in four columns, the total number of articles listed between 2009 and April 2018, the disciplines in which the Gameplay Bricks-related works were published, the total number of authors involved and their nationality, and finally their position with regard to the model. Table 4also presents the results in percentage terms. Overall, the model's diffusion is international, with France as the main country accounting for 20%. The main discipline to use the model is computer science (49%), followed by educational science (17%), art (12%), technology (12%) and CIS 10%, Critical feedback on the model accounts for the smallest percentage, 19%, behind 23.5% appropriations and a large majority of authors remaining neutral at 55.5%.

Distribution of critical positions and appropriations

Based on the data presented in Tables 2 and 3, Table 4 has been constructed to provide a more detailed breakdown of critical and appropriation stances regarding the Gameplay Bricks model. At this stage, the neutral stance has been excluded, as it does not enable the evaluation of the model. Table 3 reveals that authors from ten countries have adopted the Gameplay Bricks model, with over half of these countries being European. Conversely, authors from seven countries, more than half of which are also European, have expressed critical views of the model. From a disciplinary perspective, Communication and Information Sciences (CIS) emerges as the field with the highest level of appropriation, accounting for 30%. In contrast, Computer Science leads in terms of critical perspectives, with a rate of 50%. These findings now call for a closer examination of the nature of both appropriation and criticism, in order to rigorously evaluate the Gameplay Bricks model.

It is now time to see whether, on the one hand, other types of appropriation could be identified and, on the other, whether the set could give rise to an evaluative basis for situating its contribution to the Research.

Table 8 provides an overview of critical citations related to game brick models between 2008 and 2018. It highlights the multidisciplinary nature of research on this topic, spanning fields such as computer science, philosophy, design research and art.

This analysis reveals that computer science has the highest number of citations, reflecting its central role in the development and application of game brick models. These citations come from several countries, including Spain, Japan, the UK and the USA. We also note that most of the contributions in this field are in English, underlining the predominance of English as the main language for disseminating research on game brick models. This study demonstrates the broader theoretical and creative implications of game brick patterns with citations from philosophy, design research and art. Philosophy-related citations come notably from the USA, while design research is represented by an Australian study and art-related studies come from Canada and the Netherlands. Interestingly, while most of these publications are in English, one citation in the art category is in French, highlighting a certain linguistic diversity in the field.

In 2018, the different types of criticism identified are divided into 8 types and seem to be specific to each author: (2018, pp61-73):

Type 1: Misuse of Propp; Type 2: Subjective approach; Type 3: Lack of formalism; Type 4: Impossible classification; Type 5: Missing Meaning Bricks; Type 6: Means bricks irrelevant; Type 7: Distinguishing obligations and prohibitions; Type 8: Structure of games not studied.

In 2024, with the reading of the additional elements of the corpus, we can add a 9th type which would correspond to a formalism preventing the taking into account of storytelling or aesthetic.

Table 8. Critical Citation of Game Bricks models in the scientific literature from 2008 to 2018.

Critical Citation for Game Brick Models						
Discipline	T1-T8	Nationality	Language	Year	Nb. Aut	References
Computer Science	T3	Spain	English	2009	2	Reyno, E. M., & Cubel, J. A. (2009)
	T8	Japan	English	2010	4	Kim, T., [...] & Kondo, K. (2010)
	T9	United Kingdom	English	2015	2	Heintz, S., & Law, E. L. C. (2015)
	T4	USA	English	2015	5	Parkkila, J. [. . .] & Radulovic, F. (2015)
Philosophy	T4+T9	USA	English	2012	1	Thomas, L. D. (2012)
SIC	T2	USA	English	2008	1	Pennell, B. B. (2008)
	T5+T7	France	French	2011	1	Hurel, P. Y. (2011)
Design Research	T4	Australia	English	2017	2	Goddard W. & Muscat, A. (2017)
Art	T1+T4	Canada	French	2011	1	Arsenault, D. (2011)
	T6	Netherlands	English	2011	1	Veugen, J. I. L. (2011)
		Canada	English	2017	1	Therrien, C. (2017)

Table 9 provides an overview of critical citations of Game Brick models across various academic disciplines between 2008 and 2018. These citations indicate an analytical or evaluative engagement with Game Brick models rather than neutral references. The majority of critical citations appear in Computer Science, with six publications from diverse national backgrounds, including France, Morocco, the USA, the Netherlands, and Germany. The linguistic diversity of these citations is also notable, with publications in English, French, and German, reflecting the global discourse surrounding Game Brick models. Beyond Computer Science, critical assessments of these models are present in Science of Information and Communication (SIC) (Taiwan, 2013), History (Germany, 2011), Management (France, 2012), Education (South Korea, 2013), and Economy (Sweden, 2009). These publications are written in English, French, Korean, and Swedish, underscoring the multilingual engagement with Game Brick models in academic research. The number of authors per publication varies, from single-authored works to multi-author collaborations, suggesting different approaches to critical analysis across disciplines. The temporal distribution of these citations highlights key years of critical engagement, particularly in 2009, 2013, and 2015, indicating sustained but irregular scrutiny of the models. The presence of critical citations across multiple fields demonstrates the interdisciplinary impact of Game Brick models, with researchers actively assessing their theoretical, methodological, and practical implications.

Table 9. Neutral Citation of Game Bricks models in the scientific literature from 2008 to 2018.

Critical Citation for Game Brick Models					
Discipline	Nationality	Language	Year	Nb. Aut	References
Computer Science	France	English	2009	3	Carron, T. [...] & Mangeot, M. (2009)
	France	French	2010	1	Muratet, M. (2010)
	Morocco	English	2014	2	El Borji, Y., & Khaldi, M. (2014)
	USA	English	2015	2	Schatz, K., & Riippel, U. (2015)
	Netherland	English	2016	1	Carvalho, B., M. (2016)
	Germany	Deutch	2017	1	Piepr, J. (2017)
SIC	Taiwan	English	2013	4	Yang, H. T., [...] & Chen, K. T. (2013)
History	Germany	English	2011	1	Goelz, C. (2011)
Management	France	French	2012	3	Chollet, A., [...] & Rodhain, F. (2012)
Education	South Korea	Korean	2013	2	Kwon, C. S., Woo, T. (2013)
Economy	Sweden	Swedish	2009	1	Ahmet, Z. (2009)

Table 10 presents an overview of appropriation citations of Game Brick models across multiple academic disciplines from 2008 to 2018, classified into five subcategories (T1–T5). Appropriation citations indicate instances where researchers have integrated, adapted, or extended the Game Brick models within their work rather than merely analyzing or critiquing them. The dataset spans a wide range of fields, including Computer Science, Philosophy, Health, Management, Science of Information and Communication (SIC), Education, Language Sciences, Architecture, Design Research, and Art. The Computer Science domain exhibits the highest number of appropriation citations, with contributions from Estonia, the United Kingdom, Brazil, Italy, Germany, and Sweden, predominantly in English and Portuguese. The temporal distribution highlights increased adoption in 2014, 2015, 2017, and 2018, with author teams ranging from single to multi-author collaborations (up to eight contributors per study). This suggests a progressive incorporation of Game Brick models into computational frameworks and technological innovations. Beyond Computer Science, Philosophy (Portugal, 2016) and Health (France, 2012 and 2016) display instances of appropriation, primarily in English and French, focusing on conceptual and applied methodologies. Management (Germany, 2011) also features an English-language appropriation citation, reflecting its relevance in organizational and strategic domains. The field of Science of Information and Communication (SIC) includes citations from Mexico, Belgium, and Denmark (2009–2015), highlighting a multilingual engagement (English and French) and a growing interest in the theoretical adaptation of Game Brick models. Similarly, Education (Germany, Belgium, Russia, 2015–2016) demonstrates a diverse linguistic profile (English, French, and Russian), emphasizing the use of Game Brick models in pedagogical and instructional design. Other disciplines, including Language Sciences (France, 2016), Architecture (Turkey, 2013), and Design Research (Singapore, 2013), show targeted appropriation, indicating the versatility of these models across different research fields. Finally, Art (France, Sweden, 2011–2014) exhibits an engagement with both theoretical and applied perspectives, reinforcing the interdisciplinary impact of Game Brick models. The temporal distribution of appropriation citations reveals a steady adoption pattern, with peaks in 2014, 2015, and 2016, reflecting a maturing research interest in integrating Game Brick models into diverse disciplinary frameworks. The presence of multilingual publications and global contributions underscores the broad academic reception and adaptability of Game Brick models, reinforcing their significance as a foundational tool in various research fields.

Table 10. Appropriation Citation of Game Bricks models in the scientific literature from 2008 to 2018 with subdivision T1 to T5.

Appropriation Citation for Game Brick Models						
Discipline	T1 - T5	Nationality	Language	Year	Nb. Aut	References
Computer Science	T1 T3 T3 T1 T3 T1 T5 T1	Estonia United Kingdom Brazil Brazil Italy Germany Brazil Sweden	English English Portugese Portugese English English English English	2010 20122014 2014 2015 2017 2018 2018	1 3 1 4 8 3 3 1	Henno, J. (2010) Carter, C., [...] & Hartley, T. (2012) Murakami, L. C. […] & Almeida Macedo, D. (2014) Carvalho, B., M. , [...] & Rauterberg, M. (2015) Schmidt, S., [...] & Möller, S. (2017) Dominguez, R.G. [...] & Oliviera Venâncio, R.D. (2018) Laine, T. H. (2018)
Philosophy	T2 T2	Portugal Portugal	English English	2016 2016	1 2	Cardoso, P. J. C. (2016) Cardoso, P. & Carvalhais, M. (2016)
Health	T1 T1	France France	French French	2012 2016	3 1	Mader, S. […] & Levieux, G. (2012) Ben-Sandoun, G. (2016)
Management	T1	Germany	English	2011	4	Duin, H. […] & Thoben, K-D. (2011)
SIC	T1 T3	Mexico Belgium	English French	2009 2011	5 1	Delgado-Mata, C., [...] & Ibanez-

	T2 T2	Belgium Denmark	French English	2012 2015	1 1	Martinez, J. (2009) Hurel, P. Y. (2011) Palmieri, J. (2012) Otzen, T. (2015)
Education	T1 T4 T3	Germany Belgium Russia	English French Russian	2015 2016 2016	3 1 3	Müller, B. C., Reise, C., & Seliger, G. (2015) Gilson, G. (2016) Akchelov E.O.[...] & Nikitina K.S. (2016)
Language Sciences	T5	France	French	2016	1	Schmoll, L. (2016)
Architecture	T1	Turkey	English	2013	1	Örnek, M.A. (2013)
Design Research	T1	Singapore	English	2013	2	Yen C.C. & Lee J.M. (2013)
Art	T2 T1 T4	France France Sweden	French French English	2011 2014 2014	1 1 1	Coville, M. (2011) Fernandez, M.M. (2014) Ghys, K. (2014)

Disciplines, Citation Types, Languages, and Countries

We provide an overview of citation contexts that are not in English, highlighting their linguistic diversity and their relevance to the research. The annotation process for sentences containing citation contexts is detailed in the Table 6, where these contexts are categorized by language and corresponding annotations. Additionally, we include the translations employed during the labeling process to ensure consistency and accuracy across languages. This approach allows us to illustrate the multilingual nature of citation contexts while maintaining a standardized framework for analysis and interpretation.

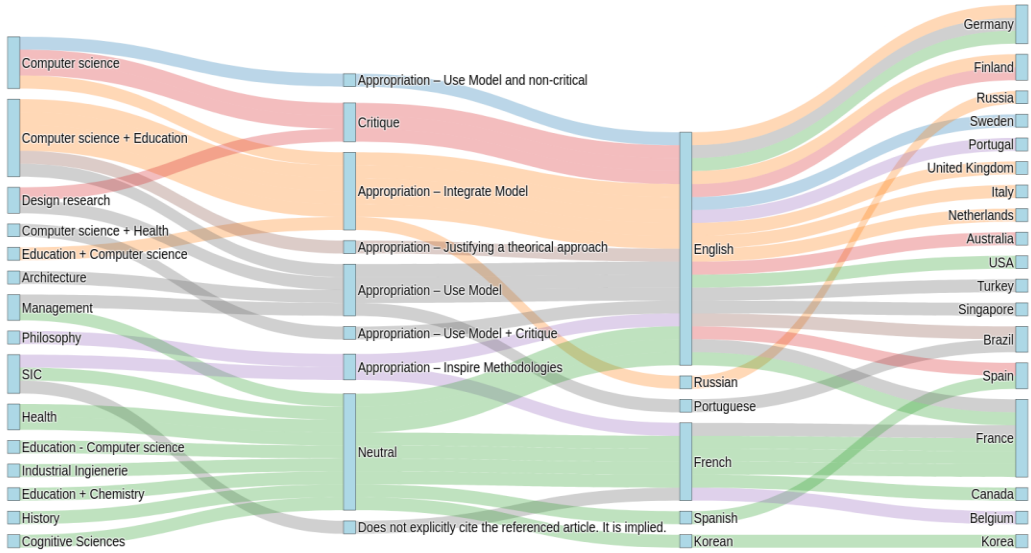

Figure 1. Citation Flow of Game Bricks Models Across Disciplines, Citation Types, Languages, and Countries.

Categories generated by the prompting approach

Figure 2 illustrates the distribution of citation context typologies within the Game Bricks research corpus during the 2008–2018 period. The data highlights the predominance of certain typologies, indicating recurring conceptual frameworks in the field, as suggested by the "Definition" and "Appropriation" categories. The dominant paradigm is thus definition and appropriation. Neutrality and criticism are more difficult to capture with our approach based on the produced sample.

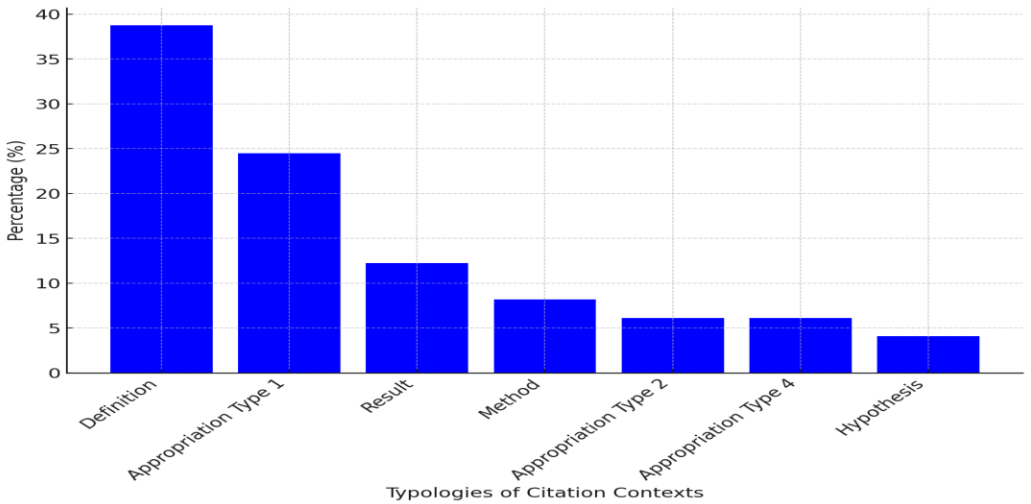

Figure 2. Distribution of Citation Context Typologies in Game Bricks Model (2008-2018).

Discussion of Experimental Results

Limitations of Our Study

The first challenge lies in the consolidation of the dataset, with a major constraint related to its multilingual dimension. Current tools do not yet offer a multilingual approach for corpus processing. As a result, its consolidation relies on human-based and time-consuming methods. For instance, Persian, Thai, and even Korean corpora could not undergo the segmentation stage, which is crucial for generating the attributes that link citation contexts to bibliographic references. The second limitation concerns the processing and assignment of attributes to citations within the text. We observed that the attributes used to associate citation segments with references are often incorrect. While this does not prevent annotation—since the citation context segment is extracted—it does hinder the ability to accurately link it to references. Moreover, handling multiple references remains challenging for this type of processing. The third limitation is the lack of adherence to citation standards in some papers, leading to processing errors. The fourth difficulty concerns the design of prompts and the reproducibility of results.

Perspectives

Despite these challenges, the approach using LLMs and prompts remains promising, provided that we can generate prompts based on categories that eliminate any semantic or conceptual indeterminacy. This is likely the next step in improving results with this type of approach. During this study, it was interesting to allow the system to propose multiple annotations for a given citation context. Granting this flexibility enabled broader coverage and improved system-generated annotations. We will focus on new reasoning models, with a particular emphasis on Chain of Thought approaches, which yielded promising results in our experiments. Indeed, the *Chain of Thought* approach will enable the explicit structuring of reasoning by breaking down a task into several intermediate steps. In citation analysis, this will allow for a better distinction between the different functions of a citation, especially in cases where citation contexts may be ambiguous. Finally, stabilizing our input corpus will allow us to conduct an evaluation comparing AI-based annotation with human annotators using the Kappa coefficient. Finally, the stabilization of our input corpus will enable us to perform an evaluation comparing AI-based annotation with human annotators using the Kappa coefficient. To this end, we will compare several llm's using tools such as LMStudio.

Conclusion

This study presents an in-depth analysis of citation contexts surrounding the gameplay bricks model between 2008 and 2018, comparing human expert analysis with AI-assisted approaches. Our results highlight both the potential and limitations of AI-assisted citation context analysis, thus emphasizing the need for hybrid approaches that integrate human expertise with machine learning capabilities.

One of the main findings of this study is the predominance of definition and appropriation categories across different disciplines, illustrating the widespread

adoption of the Gameplay Bricks model. The data reveals that computer science fields tend to appropriate this model for practical applications, while humanities and social sciences engage with it more critically. These variations highlight the influence of disciplinary conventions on citation practices and suggest that citation contexts are shaped by epistemic cultures that determine how knowledge is referenced, criticized, and integrated. Our analysis reveals the international and multidisciplinary impact of the gameplay bricks model, with citations spanning nine languages and multiple academic fields. Computer science emerges as the primary field of application (48.9%), followed by education (17.0%) and arts (12.8%), thus demonstrating the model's broad relevance. Temporal analysis shows adoption peaks in 2015-2016, suggesting a maturation phase in the model's development and application.

A second aspect concerns our methodological approach, which combines human annotation and AI-assisted classification through prompt engineering, highlighting the potential for large-scale automated citation analysis. ChatGPT-generated analyses offer advantages in terms of scalability and efficiency, enabling the processing of extended multilingual corpora that would be very time-consuming for human annotators. However, AI's ability to capture nuanced critiques and neutral citations remains limited. This limitation becomes even more pronounced when considering the expert-driven categorization of inherent critiques of Game Brick models. A detailed analysis reveals eight distinct types of criticism. Type 1 critiques highlight the erroneous application of Propp's framework (2018, p. 58), where studies misinterpret or misapply narrative structures. Type 2 critiques address a subjective approach (2018, p. 60), pointing out a lack of methodological rigor and an overreliance on interpretation. Type 3 critiques emphasize a lack of formalism (2018, p. 62), indicating that some applications fail to adopt a structured theoretical framework. Type 4 critiques argue that the model leads to an impossible classification (2018, p. 63), suggesting that its structure does not allow for a coherent categorization of game elements.

Further critiques focus on the content of Game Brick models. Type 5 critiques identify missing "Means Bricks" (2018, p. 65), arguing that essential intermediary elements necessary for game mechanics are absent. Conversely, Type 6 critiques question the relevance of certain "Means Bricks" (2018, p. 65), indicating that some components do not meaningfully contribute to game design. Type 7 critiques stress the need to differentiate obligations from prohibitions (2018, p. 66), underscoring a conceptual gap in distinguishing required actions from restricted ones. Finally, Type 8 critiques highlight the lack of analysis of game structures (2018, p. 67), pointing to a broader limitation in addressing overarching game frameworks. These identified critique categories offer a more nuanced and structured understanding of the scientific discourse surrounding Game Brick models. They emphasize not only theoretical and methodological gaps but also practical issues in the application of the framework, underscoring the need for further refinement and conceptual clarity.

This limitation results from both model biases and the inherent complexity of interpreting citation contexts, which often require deep domain expertise and understanding of implicit rhetorical subtleties. This finding aligns with previous

research on LLM capabilities in academic discourse analysis. A notable limitation of our study lies in the multilingual nature of the dataset. Current AI tools, including ChatGPT, still struggle with complex linguistic variations, particularly for underrepresented languages. While citation contexts in English and French were processed with relatively high accuracy, languages such as Persian, Thai, and Korean posed challenges due to insufficient training data and segmentation difficulties. Future research should focus on refining multilingual NLP models to better capture citation contexts across various linguistic environments. Furthermore, the gaps between human and AI-generated annotations highlight the need for more robust prompting strategies. Our results indicate that few-shot learning and chain-of-thought approaches improve AI citation classification accuracy but still cannot fully replicate human interpretative capabilities. The observed inconsistencies suggest that prompt refinement is essential for optimizing AI performance in citation analysis. The methodological challenges encountered, particularly in multilingual processing and prompt engineering, highlight important areas for future research, including:

- Developing more robust tools for multilingual citation context processing
- Improving reference linking accuracy in complex citation networks
- Refining prompt engineering techniques for specialized academic discourse
- Creating standardized evaluation frameworks for citation context analysis

In conclusion, this research contributes to the debate on AI-assisted citation analysis by proposing a comparative study spanning multiple languages and disciplines. Based on a case study, we have produced a corpus of citation contexts related to the Gameplay Bricks framework, along with prompts to categorize these contexts. We also provide a dataset of contexts annotated by an expert. Additionally, we propose a methodology for implementing categorization through prompts. It illuminates both the opportunities and challenges associated with using AI to interpret citation contexts, advocating for more sophisticated tools capable of accounting for linguistic and disciplinary variations. Moving forward, the development of improved multilingual NLP models and refinement of AI citation categorization techniques will be essential for enhancing the reliability and applicability of citation context analysis in academic research.

Acknowledgments

This work was supported by ANR-20-CE38-0003-01. The authors declare that they have no conflict of interest. This manuscript has been prepared with the assistance of artificial intelligence (AI)-based tools to support the writing process. Specifically, AI was employed for language refinement, grammar correction, and structuring of certain sections. However, all intellectual contributions, critical analysis, and interpretations presented in this work remain the sole responsibility of the authors.

References

Aithal, P., & Aithal, S. (2023). Application of chatgpt in higher education and research–a futuristic analysis. International Journal of Applied Engineering and Management Letters (IJAEML), 7 (3), 168–194.

Alvarez, J., Djaouti, D., Ghassempouri, R., Jessel, J.-P., Methel, G. (2006). Morphological study of the video games. Actes du colloque "cgie 2006". Perth, Australie.

Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 student session* (pp. 81-87).

Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.

Bertin, M., & Atanassova, I. (2024). Linguistic perspectives in deciphering citation function classification. *Scientometrics*, 1-13.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? a review of studies on citing behavior. Journal of documentation, 64 (1), 45–80.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., . . . Amodei, D. (2020a). Language models are few-shot learners. H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), Advances in neural information processing systems (Vol. 33, pp. 1877–1901).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Cabanac, G., Labbé, C., Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. evidence of critical issues affecting established journals. arXiv preprint

Cabanac, G., Labbé, C., Magazinov, A. (2022). Bosom peril" is not "breast cancer": How weird computer-generated phrases help researchers find scientific publishing fraud. Bulletin of the Atomic Scientists, January, 13.

Cabanac, G., Labbé, C., Magazinov, A. (2022). The 'problematic paper screener' automatically selects suspect publications for post-publication (re)assessment. Retrieved from https://arxiv.org/abs/2210.04895

Cardenas, Y.G., et al. (2014). Modelo de ontologia para representação de jogos digitais de disseminação do conhecimento (Unpublished doctoral dissertation). Orientador: João Bosco da Mota Alves. Coorientador: Denilson Sell. Dissertação (mestrado)-Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis.

Chen, Q., Sun, H., Liu, H., Jiang, Y., Ran, T., Jin, X., . . . Chen, H. (2023). A comprehensive benchmark study on biomedical text generation and mining with chatgpt. bioRxiv .

Coville, M. (2011). Hello, i'm a woman gamer. i got tired of people telling me to get a life. i began studying video games instead. Bruxelles, Belgique.

Delgado-Mata, C., Ruvalcaba-Manzano, R., Quezada-Patino, O., Gomez-Pimentel, D., Ibanez-Martinez, J. (2009). Low cost video game technology to measure and improve motor skills in children. Africon, 2009. africon'09 (pp. 1–6). IEEE.

Dong, C., & Schäfer, U. (2011). Ensemble-style self-training on citation classification. In *Proceedings of 5th international joint conference on natural language processing* (pp. 623-631).

Else, H. (2021). Tortured phrases' give away fabricated. Nature, 596, 328–9. Gilson, G., Draelants, H., Jardon, D., Servais, O. (2016). L'expérience virtuelle des joueurs comme situation d'apprentissage informel. Université de Mons, Belgique.
Hernández-Alvarez, M., & Gomez, J.M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. Natural Language Engineering, 22 (3), 327–349.
Hurel, P.Y. (2011). Analyse idéologique des jeux vidéo/une méthode ludonarrative pour les jeux mis en récits (Unpublished doctoral dissertation). Université de Liège, Liège, Belgique.
Jha, R., Jbara, A.-A., Qazvinian, V., Radev, D.R. (2017). Nlp-driven citation analysis for scientometrics. Natural Language Engineering, 23 (1), 93130.
Karan, B., & Angadi, G. (2023). Potential risks of artificial intelligence integration into school education: A systematic review. Bulletin of Science, Technology & Society, 43 (3-4), 67–85.
Khan, N.D., Khan, J.A., Li, J., Ullah, T., Zhao, Q. (2025). Leveraging large language model chatgpt for enhanced understanding of end-user emotions in social media feedbacks. Expert Systems with Applications, 261, 125524. https://doi.org/10.1016/j.eswa.2024.125524
Lahiri, A., Sanyal, D.K., Mukherjee, I. (2023). Citeprompt: using prompts to identify citation intent in scientific papers. 2023 acm/ieee joint conference on digital libraries (jcdl) (pp. 51–55).
Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1-35.
Lyu, D., Ruan, X., Xie, J., Cheng, Y. (2021, April). The classification of citing motivations: a meta-synthesis. Scientometrics, 126 (4), 3243–3264.
Martel, E., Lentschat, M., Labbé, C. (2024). Detection of tortured phrases in scientific literature. arXiv preprint arXiv:2402.03370
Mathebula, M., Modupe, A., Marivate, V. (2024). Chatgpt as a text annotation tool to evaluate sentiment analysis on south african financial institutions. IEEE Access.
Nishikawa, K., & Koshiba, H. (2024). Exploring the applicability of large language models to citation context analysis. *Scientometrics*, *129*(11), 6751-6777.
Radford, A. (2018). Improving language understanding by generative pretraining.
Tahamtan, I., & Bornmann, L. (2022). The Social Systems Citation Theory (SSCT): A proposal to use the social systems theory for conceptualizing publications and their citations links. *Profesional de la información*, *31*(4).
Teufel S, Carletta J, Moens M (1999) An annotation scheme for discourse-level argumentation in research articles. In: Henry S. Thompson AL (eds) Proceedings of the ninth conference on European chapter of the association for computational linguistics, Bergen, Norway, 08–12 June 1999. Association for Computational Linguistics, pp 110–117.
Teufel S, Siddharthan A, Tidhar D (2006) Automatic classification of citation function. In: Mirella Lapata HTN (editor) Proceedings of the 2006 conference on empirical methods in natural language processing, Sydney, Australia, 25–27 October 2006. Association for Computational Linguistics, 1610091, pp 103–110
Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Zhang, B., Fu, X., Ding, D., Huang, H., Dai, G., Yin, N., ... & Jing, L. (2023). Investigating chain-of-thought with chatgpt for stance detection on social media. *arXiv preprint arXiv:2304.03087*.

Zhang, Y., Wang, Y., Wang, K., Sheng, Q. Z., Yao, L., Mahmood, A., ... & Zhao, R. (2023). When Large Language Models Meet Citation: A Survey. *arXiv preprint arXiv:2309.09727*.