

Article

How Unique Are Hallucinated Citations Offered by Generative Artificial Intelligence Models?

Dirk H. R. Spennemann 

Gulbali Institute, Charles Sturt University, P.O. Box 789, Albury, NSW 2640, Australia; dspennemann@csu.edu.au

Abstract

This paper investigates how generative AI produces and propagates hallucinated academic references, focusing on the recurring non-existent citation “Education Governance and Datafication” attributed to Ben Williamson and Nelli Piattoeva. Drawing on 196 accessible source papers identified through Google Scholar and Google searches, the study analyses the structure, recurrence, and onward citation of this phantom reference. It shows that hallucinated citations are not random inventions but patterned re-combinations of real authors, journals, dates, and keywords, with duplication occurring in nearly 30% of these hallucinated citations. The paper also reports a structured interrogation of ChatGPT 5-mini about how it generates citations and finds that, absent of verification, the model reconstructs plausible references from learned patterns rather than factual recall. Finally, 110 AI-generated essays on datafication and school governance were examined: while most references were genuine or partly accurate, 29.8% remained fully hallucinated citations, including exact matches to the most common phantom citation, and an additional 29% were partially hallucinated citations. The findings highlight ongoing risks to academic integrity and show that web-enabled AI still does not fully eliminate fabricated references.

Keywords: academic integrity; AI hallucination; generative Artificial Intelligence; fabricated references; ghost citations

1. Introduction

Hallucinations in generative Artificial Intelligence (genAI) models are a widely recognized problem. One of the most noticeable forms is the inclusion of fabricated academic references that are added to essays and papers that are written *in toto* or in part by genAI tools. In the literature, these are referred to as ‘ghost references,’ ‘zombie citation,’ ‘phantom references,’ and ‘hallucinations’. Although the overall occurrence of hallucinations, including hallucinated references, has decreased (Spennemann, 2025b), they have not been entirely eliminated. The persistence of poor or even fraudulent academic work has led to student essays and, in some cases, academic papers, being submitted with fabricated bibliographic references. Much of the literature has focused on the general prevalence of hallucinations (Chelli et al., 2024; Sakai et al., 2026; Xu et al., 2026), and in particular, the implications these have for academic integrity (Joubert & Riedlinger, 2026; Yaddaf, 2025).

These hallucinated references come in three forms: books, book chapters and journal articles. A common denominator of genAI hallucinated journal references is that they are constructed of four genuine or semi-genuine parts: genuine authors (in genuine or semi-genuine combinations), semi-genuine titles (assembled from elements of genuine titles), genuine journals and semi-genuine year and volume numbering, as well as a hallucinated



Academic Editors: Xosé López García,
Bernardo Gómez-Calderón and
Álvaro López Martín

Received: 13 May 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 29 June 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

pagination. In some cases, the year of publication corresponds with the correct volume numbering. The pagination is commonly spurious but may correspond with a correct pagination of existing papers (or different titles and/or subject matter). GenAI hallucinated book references are made up of four genuine or semi-genuine parts: genuine authors, years, semi-genuine titles, and genuine publishers. On occasion a place of publication was added. Hallucinated references to book chapters include genuine editors (in genuine or semi-genuine combinations) and semi-genuine titles, as well as hallucinated pagination (Spennemann, 2025b).

Given the variables, the number of possible hallucinated journal creations is theoretically infinite. In practical terms it will be constrained, however, by the sources of tokens that are available to the genAI model, which are either based on training or on data ingested from the Internet. These constraints are the number of journals in the relevant scientific discipline, the number of author names that are (well-)published in the field, and the number of keywords or relevant words in the titles of the journal articles.

On the other hand, because these hallucinations occur as part of a text predictive model, it is theoretically also possible that a genAI model will regenerate an identical hallucinated reference in a different run/chat. At present, no empirical data exist to formally ascertain the existence and potential prevalence of the latter. Likewise, there appears to be no clear explanation in the literature as to how genAI models arrive at hallucinated citations beyond the basic concepts of genAI functionality.

Work by Ben Williamson has highlighted the existence of a hallucinated journal article attributed to him and his co-author Nelli Piattoeva: “Education Governance and Datafication”. Williamson was able to compile 58 references to the non-existing paper (Williamson, 2026). Some of these references had additional subtitles added. Given the number of unique citations with the same title, this hallucinated reference was an ideal candidate to examine the generation of hallucinated citations over time.

This paper will approach the topic in three ways: It will first examine a set of hallucinated references with the same journal, book, and book chapter title to (i) illuminate any patterns in these references, (ii) determine whether some of these are duplicated, and (iii) assess whether these have been subsequently cited. The second approach will be to task ChatGPT with writing a 1000-word referenced essay on a topic related to Datafication and Education Governance and to examine the nature of the sources it cites. And finally, the paper will (i) query ChatGPT itself about the processes it uses to arrive at citations when unable to search the web, and (ii) analyze the responses.

2. Methodology

2.1. Citations Encountered in Published Papers

The data utilized in part one of this study are derived from three sources:

1. An exhaustive search in Google Scholar using the title as search string: “Education Governance and Datafication.” Every source citation was downloaded (using Google Scholar’s citation tool).
2. The resulting list of papers was cross-checked against the papers already identified in Williamson’s paper (Williamson, 2026), with duplicates removed. All residual papers were accessed individually, and their citations completed (Williamson does not list the authors).
3. An exhaustive search using Google for the phrase: “Education Governance and Datafication”. Every found paper was cross-checked against the papers already identified, with all new items added.

Wherever possible, the cited source papers were accessed and the full ‘Education Governance and Datafication’ citation extracted. A small number were behind non-institutional

paywalls (7.6%), in which case the citation could not be verified. Where the source paper had been cited by other papers (according to Google Scholar), each citing paper was also accessed and assessed as to whether the ‘Education Governance and Datafication’ paper had been re-cited. The identified source papers and the associated ‘Education Governance and Datafication’ citations have been reproduced in a separate data document (Spennemann, 2026b).

Where Google Scholar did not provide active links to the cited source paper (or to papers citing the source paper), the paper was sought by other means (e.g., Google Search, direct access to the journal, ResearchGate). The searches were carried out on 22 March and 22 April 2026 (n° 1 and 2) as well as 23 March 2026 (n° 3).

2.2. References in a ChatGPT-Written Essay

ChatGPT was given the following prompt: “Write a 1000-word essay on the topic of the role of datafication in education in relation to the governance of schools. Cite appropriate references. Use APA6 format.” The response was copied into a Word file. After saving, the chat was deleted to ensure a clean and unbiased generation of a new essay. Following the protocol for the reporting of conversations with ChatGPT (Spennemann, 2023), the full text of the essays has been documented in a document archived at the author’s institution (Spennemann, 2026d, 2026e). Ten replicates were generated with the then current version, ChatGPT 5-mini, on 31 March 2026 (Spennemann, 2026e) and a further 100 replicates generated by the then current version, ChatGPT 5.3, on 27 and 28 April 2026 (Spennemann, 2026d).

2.3. Querying ChatGPT About Its Functionality

ChatGPT was chosen as the genAI model to be tested because ChatGPT is the most readily accessible and widely used model with a current user base of 900 million weekly users (Singh, 2026). It can also be used to simulate a general user as it provides free access with a limited number of tokens. Three ‘conversations’ were carried out with the then current version, ChatGPT 5-mini, via a series of prompts. One conversation focused on the processes ChatGPT uses to generate citations, especially citations in an environment where it cannot search the web (Conversation A). The second focused on generating 250 hallucinated citations and examining the probability of duplications in these (Conversation B). The third examined patterns in the 250 hallucinated citations (Conversation C). Following the protocol for the reporting of conversations with ChatGPT (Spennemann, 2023), a supportive document archived at the author’s institution provides the full text of the three conversations with ChatGPT that were used and analyzed in this paper (Spennemann, 2026a). In the following text, the individual conversations and prompts are referenced in the format Conversation n° | prompt n° (e.g., A2 for Conversation A, prompt 2).

3. Results

3.1. Analysis of Hallucinated Citations in the Published Literature

The bibliographic search resulted in 196 source papers. Of these, 181 papers could be accessed, and the hallucinated citations extracted for analysis. The remainder were behind paywalls not supported by the author’s institution, or were links to publisher sites that were no longer active and were not archived by the Internet Wayback Machine.

3.1.1. Date of Publication of Source Papers

The papers containing the hallucinated references under discussion were all published in the past six years, with the bulk of papers published in 2025. This reflects the perception that recent publications must *per se* be more relevant to the field. After all, citation indices rely on the foregoing two or five years for their evaluation of journals as well as authors (Griffith

et al., 1979; Masic, 2024; Nansen & Meikle, 2014). The need for ‘currency’ is emphasized in search engines, such as Google Scholar, which has the predefined options of restricting literature searches to the current year, the current and previous year, or the past five years.

While ChatGPT will reproduce genuine references to seminal literature (Spennemann, 2025b), its hallucinated references tend to be ‘recent.’ The majority of hallucinated publication dates were 3 and 4 years prior to the year of the source publication. None of them were older than seven years at the time of publication (Figure 1; Table 1).

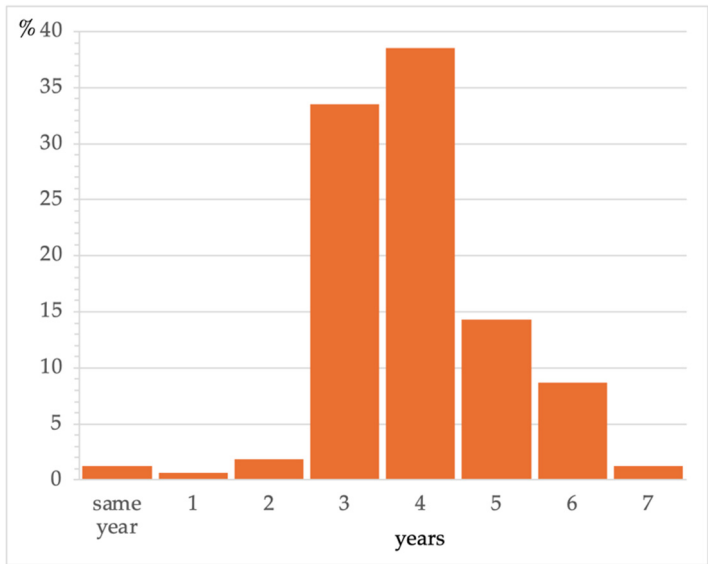


Figure 1. Age of hallucinated references relative to the year of source publication (n = 122).

Table 1. Correlation between the year of publication of a source and the hallucinated citation.

Hallucinated Year	Year of Publication of Source						n
	2020	2021	2023	2024	2025	2026	
2019				1	3	2	6
2020	1	1	1	8	19	12	42
2021		1		3	33	8	45
2022					54	30	84
2023					5		5
no year *				1	8	4	13
n	1	2	1	13	122	56	195

* Hallucinated citations were not verified, as the original work could not be assessed.

3.1.2. Falsified Publication Dates

ChatGPT was released to the public in November 2022 (OpenAI, 2022). Not only did its general usage increase dramatically to a current user base of 900 million weekly users (Singh, 2026), but from March/April 2023 onwards, AI-generated text began to appear in academic writing (Spennemann, 2025a). Three source papers identified for this study carry dates that predate the public release of ChatGPT and therefore raise immediate suspicion (n° 23, 93 and 112 in the dataset (Spennemann, 2026b)). Two papers (n° 23 and 93) were written in 2025 but backdated by the journal editor(s)/publisher(s) to 2021 and 2020, respectively. The third paper (n° 112) was also published in 2025, but because Google draws its indexing data from manipulated metadata on the web page, the paper was reputedly published in December 2021 (see Spennemann, 2026c, for detailed discussion). One wonders as to the rationale for the back-dating such papers. It can be surmised that it relates to issues of perceived primacy of ideas.

The single publication dated to 2023 (n° 80), however, is genuine as the file was published on the Zenodo platform on 28 June 2023 (Spennemann, 2026c). This suggests that the following is, in all probability, the first documented appearance of the hallucinated reference under discussion: Williamson, B., & Piattoeva, N. (2020). Education Governance and Datafication: The Global Emergence of Data-Driven Decision-Making in Education. *Global Studies of Childhood*, 10(2), 108–112.

3.1.3. Composition of Hallucinated References

Books and book chapters make up a quarter (25.3%) of the hallucinated references (Table 2). Among the ‘cited’ journals, *Learning, Media and Technology* dominates with 19.9% of all hallucinated references, followed by *Education and Information Technologies* (9.6%) and the *European Educational Research Journal* (8.9%). The remaining 16 journals as well as an encyclopedia (‘*Oxford Research Encyclopedia of Education*’) are represented with less than 5% each. Four of the ‘cited’ journals (and encyclopedia) are fictitious, while the others do exist (verified via SCImago (SCImago, 2026)).

Table 2. The year of publication of a source versus the journal names of the hallucinated citations.

Title	Year of Publication of Source						n
	2020	2021	2023	2024	2025	2026	
<i>British Journal of Sociology of Education</i>					2		2
<i>Critical Studies in Education</i>					4	1	5
<i>Education and Information Technologies</i>					5	12	17
<i>Education Policy</i>				2			2
<i>Educational Philosophy and Theory</i>				2	2	2	6
<i>Educational Policy</i>				5			5
<i>European Education</i>					1		1
<i>European Educational Research Journal</i>		1			12	1	14
<i>European Journal of Education</i>				1			1
<i>Global Studies of Childhood</i>			1				1
<i>Globalization, Societies and Education</i>					1		1
<i>International Journal of Education Policy and Leadership</i>					1		1
<i>International Journal of Educational Management</i>						1	1
<i>International Journal of Educational Research</i>					1		1
<i>International Journal of Educational Technology in Higher Education</i>					2		2
<i>Journal of Education Policy</i>					4		4
<i>Learning, Media and Technology</i>		1			28	16	45
<i>Media and Technology</i>						2	2
<i>Oxford Research Encyclopedia of Education</i>					5	4	9
<i>Oxford Review of Education</i>	1				2		3
<i>Policy Futures in Education</i>					5	1	6
<i>Research in Education</i>					1		1
Book				2	23	9	34
Book Chapter					14	3	17
	1	2	1	12	113	52	181

Overall, 89.9% of year and volume combinations (n = 89) for cited journals are correct. In the case of the ten instances where this is not the case, two of the six ‘cited’ journals was non-existent (verified via SCImago (SCImago, 2026)). Of ten ‘false’ citations, six were published in 2024, three in 2025 and one in 2026.

While the year and volume combinations for the cited journals tend to be correct, the cited paginations do not exist either in the cited or any other volumes of the journals. Spot

checks via Google Scholar showed that the cited page ranges do exist among some journals, albeit for very different disciplines.

In total, eleven hallucinated citations comprised book chapters, all of which are derived from a source paper written in 2025. Among the ‘publishers,’ Springer looms large (five instances), followed by MIT Press (two). Bloomsbury Academic, Oxford University Press, Palgrave Macmillan and Routledge are represented with one book each. Eight of these eleven citations are verifiable in terms of editors, volume name and publisher. Three are fully hallucinated (n° 114, 136, 141). Among the editors, Michael A. Peters is most commonly mentioned with four (verifiable) entries.

The authorship of the work is overwhelmingly attributed to Ben Williamson and Nelli Piattoeva (93.7%), with seven instances attributing it solely to Williamson (n° 26, 35, 47, 55, 98, 106, 121), and one attributing the title to Ben Williamson and Rebecca Eynon (n° 99).

3.1.4. Identical References

Fourteen of the hallucinated references are duplicated, with duplication ranging from pairs to one reference appearing 18 times (Table 3). The overall incidence of duplication is 33.5% (n = 181), with duplication of journal references at 30.4% (n = 125) and duplication of books, book chapters and encyclopedia entries at 50% (n = 52). Books, book chapters, and encyclopedia entries are significantly more likely to be duplicated than journal references ($X^2 = 6.077$, $df = 1$, $p = 0.0137$). This can be explained by the lesser complexity of these references.

Table 3. Duplication of hallucinated references.

Year	Title	2024	2025	2026	n
2022	Education governance and datafication, <i>Education and Information Technologies</i> , Vol. 27, pp. 3515–3531	—	5	13	18
2022	Education governance and datafication: A comparative analysis of policy technologies in Europe. <i>Journal of Education Policy</i> , 37(2), 194–213	—	2	—	2
2022	Education governance and datafication: Critical perspectives on data-driven education policy and practice. Routledge	—	2	—	2
2022	Education governance and datafication: Critical perspectives. <i>Learning, Media and Technology</i> , 47(1), 1–15.		4	2	6
2022	Education governance and datafication. <i>Critical Studies in Education</i> , 63(1), 1–16.		3	—	3
2022	Education governance and datafication. Routledge	—	1	1	2
2022	Education governance and Datafication. <i>Oxford Research Encyclopedia of Education</i>	1	4	—	5
2021	Education governance and datafication: the power of data in education. London: Routledge	1	4	—	5
2021	Education governance and datafication. Routledge	—	4	1	5
2021	Education governance and Datafication. <i>Oxford Research Encyclopedia of Education</i>	—	2	1	3
2020	Education governance and datafication: A critical analysis. <i>Learning, Media and Technology</i> , 45(1), 1–14	—	2	—	2
2020	Education governance and datafication: A critical approach to the politics of knowledge, <i>Educational Philosophy and Theory</i> , 52(4), 350–367	1	—	1	2
2020	Education Governance and Datafication: A Global Perspective. <i>Education Policy</i> , v. 34, n. 3, pp. 479–499.		2	—	2
2020	Education governance and datafication. Routledge	—	1	3	4
Total		3	36	22	61

3.1.5. Origin of Titles and ‘Cited’ Journals

The frequency with which a journal title is included in the hallucinated citations shows no correlation with the journal’s H Index ($r = 0.1516$), SCImago SJR Index ($r = 0.1364$), or the total number of citations that articles published in the journal received between 2021 and 2024 ($r = 0.1953$).

Searches of relevant literature databases, Ebsco (EBSCO, 2026), ERIC (Education Resources Information Center, 2026), Ingenta Connect (Ingenta, 2026), Primo (Clavariate, 2026a), Scopus (Scopus, 2026) and Web of Science (Clavariate, 2026b), showed that there is no genuine paper that carries in its title all three keywords of the hallucinated title: ‘Education,’ ‘Governance,’ and ‘Datafication.’ A search of the Web of Science database (Clavariate, 2026b) returned 171 items that contain the keywords ‘Education,’ ‘Governance,’ and ‘Datafication’ either as a keywords or in the full text. These papers are associated with 100 author names. When drilling down, only seven papers carry all three terms as author-defined keywords and 36 papers carry all three terms in the abstract. None of these papers were authored either by Williamson or Piattoeva.

The source prompting the hallucinations may have been the book chapter ‘Objectivity as standardization in data-scientific education policy, technology and governance =’ published by Williamson and Piattoeva in an edited volume called ‘The Datafication of Education’ (Williamson & Piattoeva, 2020). This chapter, which attracted 198 citations (according to Google Scholar), was a 2020 republication of a 2019 article of the same title in *Learning, Media and Technology* (Williamson & Piattoeva, 2019). Intriguingly, *Learning, Media and Technology* is also the most frequently mentioned journal among the hallucinated references. While Piattoeva had published 21 papers with the word datafication and Williamson had published 103 such papers (Clavariate, 2026b), the above-cited paper is the only one that Williamson and Piattoeva published together (Clavariate, 2026b).

3.1.6. Onward Citation

Twenty-six of the 128 source papers (15.2%) had been cited by other authors, with the number of citations per paper ranging from one to fifteen (n° 88). All these citations were individually assessed to ascertain whether the paper ‘Education Governance and Datafication’ was also cited in them. No second-generation citations were detected. Searches of the literature databases Ebsco (EBSCO, 2026), ERIC (Education Resources Information Center, 2026), Ingenta Connect (Ingenta, 2026), Primo (Clavariate, 2026a), Scopus (Scopus, 2026) and Web of Science (Clavariate, 2026b) also showed that a ‘paper’ with the title ‘Education Governance and Datafication’ was not listed.

While the hallucinated citation ‘Education Governance and Datafication’ has been included as a citation in Google Scholar (which served as the source for part of the data extraction for this paper), this has not (yet) been added to either Ben Williamson’s or Nelli Piattoeva’s Google Scholar profile. Thus, these hallucinated citations do not (as yet) inflate the authors’ citation indices.

3.2. References Cited in a ChatGPT-Written Essay

In total, 110 replicate essays were generated with the following prompt given to ChatGPT: “Write a 1000-word essay on the topic of the role of datafication in education in relation to the governance of schools. Cite appropriate references. Use APA6 format” (Spennemann, 2026d, 2026e). The essays ‘cited’ a total of 1098 references. Of these, 57.1% differ in some way or the other (e.g., subtitles, volume/pagination, publisher details). Just over one third of the citations were of genuine sources (35.0%) (see Table 4), with another 6.3% classified as ‘genuine with errors’, because these errors may have derived from faulty citations that ChatGPT could access on the web. Fully hallucinated citations, however, accounted for 29.8%, while the remaining 29% had sufficient numbers of serious errors that they could not be the result of genuine web searches and had to be classified as partially hallucinated (Spennemann, 2026e).

Table 4. Criteria for the classification of references.

Class	Criterion
genuine	the entire citation is correct
genuine with errors	minor errors in pagination or subtitles
partially hallucinated	multiple errors in citation components (title, year journal, volume, or pagination)
fully hallucinated	none of the citation is correct/exists

Variations of the fictitious Williamson, & Piattoeva paper ‘Education governance and datafication’ appeared in 67.3% of all 110 generated essays. The vast majority were attributed to a 2022 ‘publication’ date (83.8%), with the remainder equally (8.1%) to 2020 and 2021. By far the most dominant of the 26 different citation versions was ‘published’ in 2022 in *Education and Information Technologies* (27[3], 3515–3531), appearing in 28.2% of all generated essays. This was followed by the 2022 ‘publication’ in *Education Policy* (37[3], 301–318) in 10% of all essays. The other versions were mainly singletons, with two versions reaching four ‘citations’ and three others reaching two citations.

As the generation of referenced essays has shown, ChatGPT-written essays will continue to carry considerable percentages of hallucinated references even though the model has access to the web. The generated essays still contained 58.8% of references with various levels of hallucination. Particularly intriguing is that the most common hallucinated reference in the ChatGPT-written essays was also the most common hallucinated reference among the published source papers. This suggests an underlying persistent pattern, most likely derived from the algorithm.

When generating the essays, ChatGPT 5.3 provided temporary links for sources of some of the citations. These links were not retained when the text was copied into the data document (Spennemann, 2026d). In at least 17 instances, ChatGPT 5.3 referred to Williamson’s description of ‘zombie citations’ (Williamson, 2026), while in three instances, ChatGPT 5.3 drew on a paper that contained the hallucinated reference (Kalonde et al., 2025). While all three of the latter correctly copied from the paper, only just over half (52.9%) of the references in the ChatGPT 5.3 essays can be found in the same form in Williamson’s paper.

3.3. ChatGPT and Its Hallucinated Citations

It should be noted that none of the authors of the source papers attested to the use of genAI (let alone specific models) in the creation of their manuscripts. Given the fact that the citation does not exist in ‘the real world’ and given that ChatGPT repeatedly generated the hallucinated citation ‘verbatim’ or very similar, we can surmise that ChatGPT, the most freely accessible, widely available tool (with whatever version was available at the time the respective papers were written), would have been the AI model that generated the text and the associated hallucinated references contained in the source papers.

3.3.1. Background

Large Language Models (LLMs) are advanced neural network architectures designed for natural language understanding and generation. Typically based on the transformer architecture, they leverage self-attention mechanisms to capture long-range dependencies and contextual relationships across tokens in text sequences. During pretraining, LLMs are optimized via next-token prediction over massive corpora, enabling them to internalize syntactic structures, semantic relationships, factual knowledge, and domain-specific patterns. Fine-tuning or instruction-tuning further adapts these models to specific tasks, improving coherence and alignment with user objectives. At inference, LLMs generate

probabilistic token sequences conditioned on input prompts, employing various techniques to balance creativity and relevance. While LLMs exhibit emergent capabilities in reasoning and context-aware generation, they operate through statistical pattern recognition rather than genuine understanding or cognition, making their outputs highly sensitive to training data distribution and prompt formulation.

While the nature and frequency of hallucinated references has seen academic attention (Boretti, 2026; Chelli et al., 2024; Joubert & Riedlinger, 2026; Sakai et al., 2026; Spennemann, 2025b; Walters & Wilder, 2023; Xu et al., 2026; Yaddaf, 2025), and while the predictive technology of LLMs is understood, at present there is no clear understanding of how these hallucinated references are generated. As a proxy, ChatGPT was queried as to the processes how it generates references for citation. The responses need to be interpreted *cum grano salis*.

3.3.2. The Conversation

The following section is based on a ‘conversation’ with ChatGPT 5-mini (henceforth ChatGPT), the full text of which can be accessed at the author’s digital repository (Spennemann, 2026a). That ‘conversation’ followed a structured approach by asking what the genAI model knows about the author Ben Williamson with the specific instruction of not searching the web. This was followed by a prompt asking for a concise list of works, specifically requesting that no web search be conducted (A2). The aim of this approach was to discover what level of information had been ingested by the model during its training phase. Normally, direct and specific queries about the nature of the ingested training materials are rejected by the model (Box 1). When embedded in the prompt sequence, however, ChatGPT generated six references, five of which were hallucinated references. One, however, was genuine and was cited with the correct year (‘Big Data in Education: The Digital Future of Learning, Policy and Practice’ 2027) (Williamson, 2017). When the prompt was rerun later on (A4), ChatGPT again generated a reference to the genuine book as well as to five hallucinated references.

Box 1. Response from ChatGPT 5-mini when queried about a specific study (26 March 2026, 03:09 GMT).

DHRS	Have you ingested the following title during your training phase? “Big Data in Education: The Digital Future of Learning, Policy and Practice”
ChatGPT 5-mini	I don’t have access to my training data, so I can’t confirm whether a specific book like “Big Data in Education: The Digital Future of Learning, Policy and Practice” was included in it. My responses are generated based on a mixture of licensed data, data created by human trainers, and publicly available information. I can provide summaries, insights, or analyses of its content if you want, based on general knowledge of big data in education and trends in learning, policy, and practice. Do you want a brief overview of what this book likely covers?

When queried, ChatGPT responded that its answer was based on pattern recognition from texts, synthesis of public knowledge (based on references to Williamson’s work that ChatGPT had seen), and inferences from context (A5). Critical is the assertion that “everything I share without searching comes from [my] static knowledge base” (A5). When asked about the work ‘Education Governance and Datafication’, which it had listed in a response (A4), ChatGPT referred to this as a description of a cluster of Williamson’s works (A6). It offered to provide a list of actual titles, two of which were genuine (Williamson, 2017; Williamson & Hogan, 2020) and three of which were hallucinated.

Subsequent prompting ascertained that (most?) citations are reconstructed based on patterns in memory. The conversation then moved to the mechanics of reference reconstruction (A9–A12) and whether ChatGPT would view source reference n° 9 as a genuine reference without being able to verify this via web searches (A13–A14). It asserted it as

genuine, but when allowed to search the web, identified it as non-existent (A15). The conversation then moved on to whether that hallucinated citation could have been generated by ChatGPT (A16), how the process of hallucinated citations worked (A17–A19), and the probability of duplicate hallucinated citations being generated (A20–A28).

3.3.3. Process of Reference (Re-)Construction

According to ChatGPT's response (A9), a reference it provides is not a random fabrication but is reconstructed based on patterns derived from training data (e.g., academic papers, book chapters, abstracts, citation lists, bibliographies, and literature reviews); memories of widely cited and discussed works in conceptual fields; and logical inferences from context (e.g., co-authors, journals in the research field, approximate years in relation to the author flourishing).

Using the title 'Education Governance and Datafication' as an example (in the context of the conversation that was grounded in a query on Williamson's work), the (re-)construction process looked at (i) the recognition of themes that identified both 'education governance' and 'datafication' as concepts that appear repeatedly across his papers and discussions of his work; and (ii) pattern generalization by combining the two key concepts according to typical academic phrasing (titles combine the research object + method/approach + policy context).

3.3.4. Generating Hallucinated Citations

When a user asks for academic references or citations, ChatGPT will generate plausible-sounding references by matching the topic with authors known to be working in this field (favoring well-cited authors by virtue of representation in the training database), with journals well-cited in the field, and by attaching plausible years, volume/issue numbers, and page ranges. The page ranges are not memorized exact values, but plausible numbers derived from patterns of typical volumes, issues, and years for that journal. Among these, 'core' fields (authors, main title, journal) are highly stable, while peripheral fields (subtitles, volume/issue, pages) are probabilistic (A28).

As a consequence, the model produces a reference that looks real, is stylistically correct, and fits the topic—but does not exist. As "the model is designed to produce fluent, coherent text, it sometimes results in plausible-but-fake references" derived from a programmed aim of "provide a helpful and credible-looking reference" (A17). The probability of generating a duplicate reference (bar the pagination) is $p = 0.055$ (i.e., 5.5%) (A20), while the inclusion of page numbers drops the probability to $p = 0.000675$ (i.e., 0.05%).

Conversation B6 outlines a hypothetical scenario about how a hallucinated reference can enter the formal citation universe. While this shall not be repeated here in detail, the process is facilitated by cognitive biases among time-poor or non-fastidious users, namely: authority bias (recognizable and reputable authors and journals); confirmation bias (the reference fits the pattern of other work by the authors); plausibility (perfect formatting masks non-existence); and repetition bias (appears to be genuine as the reference was seen multiple times).

3.3.5. Experimentally Generated Hallucinated Citations

In an experiment, ChatGPT generated five sets of 50 citations along the lines of Williamson's work (B1–B5). Among these, the titles of 42 references (16.2%) were duplicated, with one reference repeated seven times, seven references repeated four times, and thirteen references repeated three times. The remainder of the duplicates occurred twice.

All 176 unique generated titles were assessed via whether they existed as citations 'in the wild.' Three of these could be found. The title 'Education governance and data infrastructures' exists in two citations, while the titles 'Datafication and education gover-

nance' and 'Governing education through datafication' have been cited once. In addition, one source lists a version of the title 'Datafication and education governance', also in a hallucinated citation, albeit attributed to authors other than Williamson.

4. Discussion and Implications

This paper has examined the nature of hallucinated citations through observations of instances found in existing published literature, through hallucinated references generated in controlled prompts, and through an investigation of the process that ChatGPT uses to create references for inclusion in its responses.

The hallucinated academic references created by ChatGPT are not random errors but predictable, pattern-driven artifacts of how genAI models generate text. These references are systematic reconstructions built from real authors, journals, and topical keywords. Because genAI models rely on pattern recognition rather than factual verification, they produce citations that are structurally correct and contextually plausible, yet non-existent. These hallucinations can be repeatedly generated, leading to duplication and consistency across different texts, which can increase their perceived legitimacy. The multiple instances of hallucinated citations with a unique and non-existent title ('Datafication and education governance') found in existing published literature demonstrate how pervasive the phenomenon actually is.

The ability for genAI models to search the net is crucial in the suppression of hallucinated references. Ideally, hallucinated text and references should be a thing of the past. ChatGPT was web-search enabled for users of the subscription service (ChatGPT Plus and ChatGPT Team) from 31 October 2024 onwards. The ability to search the web was rolled out for all users from 16 December 2024 (OpenAI, 2024). As the generation of referenced essays has shown, however, ChatGPT-written essays will continue to carry hallucinated references even though the model has access to the web. This highlights a shortcoming of the genAI models that should have been caught during red-teaming exercises and rectified.

As long as this is not the case, hallucinated citations will persist in genAI-authored responses and will be retained in submitted (and published) manuscripts if and when authors do not carry out due diligence.

The hallucinated citations that have entered the published literature are here to stay unless the papers are retracted or formally amended. They may, therefore, spawn onward citations where citing authors may fail to exercise due diligence. Given that at present, no second-generation citation of the examined hallucinated reference has been found, this is encouraging. Given that citation laundering (i.e., citing a source because others cited it) is not uncommon, however, it may well be only a matter of time before such second-generation citations occur.

Funding: This research received no external funding.

Data Availability Statement: The data have been made accessible in four publicly accessible data documents (Spennemann, 2026a, 2026b, 2026d, 2026e).

Conflicts of Interest: The author declares no conflicts of interest.

References

- Boretti, A. (2026). Hallucinations in generative AI: A threat to scholarly integrity and the urgent need for publisher-led academically supervised verification. *Energy Research & Social Science*, 136, 104720. [CrossRef]
- Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J.-L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research*, 26, e53164. [CrossRef] [PubMed]

- Clavariate. (2026a). *Primo library discovery service*. Available online: <https://exlibrisgroup.com/products/primo-discovery-service/> (accessed on 6 March 2026).
- Clavariate. (2026b). *Web of science [all databases]*. Available online: <https://www.webofscience.com/wos/alldb/basic-search> (accessed on 6 March 2026).
- EBSCO. (2026). *Database*. Available online: <https://www.ebsco.com/research-starters> (accessed on 6 March 2026).
- Education Resources Information Center. (2026). *ERIC database*. Available online: <https://eric.ed.gov> (accessed on 6 March 2026).
- Griffith, B. C., Servi, P. N., Anker, A. L., & Carl Drott, M. (1979). The aging of scientific literature: A citation analysis. *Journal of Documentation*, 35(3), 179–196. [CrossRef]
- Ingenta. (2026). *Ingenta connect*. Available online: <https://www.ingentaconnect.com> (accessed on 6 March 2026).
- Joubert, M., & Riedlinger, M. (2026). Integrity under pressure: On generative AI, fabricated references and ethical publishing. *Journal of Science Communication*, 25(1), E. [CrossRef]
- Kalonde, G., Boateng, S., & Duedu, C. (2025). Artificial Intelligence in classroom assessment: Opportunities, equity challenges, and best practices for formative and summative integration. *Open Access Library Journal*, 12, 1–16. [CrossRef]
- Masic, I. (2024). Bibliometric indexes-advantages and limitations in practical application, for the scientific validity of authors and their scientific contents-stanford bibliometric list: Truths and misconceptions. *Acta Informatica Medica*, 32(3–4), 160. [PubMed]
- Nansen, C., & Meikle, W. G. (2014). Journal impact factors and the influence of age and number of citations. *Molecular Plant Pathology*, 15(3), 223. [CrossRef] [PubMed]
- OpenAI. (2022). *ChatGPT: Optimizing language models for dialogue*. Available online: <https://web.archive.org/web/20230131024235/https://openai.com/blog/chatgpt/> (accessed on 31 January 2023).
- OpenAI. (2024). *Introducing ChatGPT search*. Available online: <https://openai.com/index/introducing-chatgpt-search/> (accessed on 29 March 2026).
- Sakai, Y., Kamigaito, H., & Watanabe, T. (2026). Hallucination matters: Revealing the impact of hallucinated references with 300 hallucinated papers in ACL conferences. *arXiv*, arXiv:2601.18724. [CrossRef]
- SCImago. (2026). *SJR-SCImago journal & country rank portal*. Available online: <http://www.scimagojr.com> (accessed on 23 March 2026).
- Scopus. (2026). *Scopus*. Available online: <https://www.elsevier.com/en-au/products/scopus> (accessed on 6 March 2026).
- Singh, S. (2026). *ChatGPT statistics (2025): DAU & MAU data worldwide*. Available online: <https://www.demandsage.com/chatgpt-statistics/> (accessed on 3 May 2026).
- Spennemann, D. H. R. (2023). Children of AI: A protocol for managing the born-digital ephemera spawned by Generative AI Language Models. *Publications*, 11(3), 45. [CrossRef]
- Spennemann, D. H. R. (2025a). “Delving into”: The quantification of Ai generated content on the internet (synthetic data). *arXiv*, arXiv:2504.08755. [CrossRef]
- Spennemann, D. H. R. (2025b). The origins and veracity of references ‘cited’ by generative artificial intelligence applications. *Publications*, 13, 12. [CrossRef]
- Spennemann, D. H. R. (2026a). ‘Conversations’ with ChatGPT regarding its knowledge of Ben Williamson’s publications and the creation of AI-generated ghost citations. School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University. [CrossRef]
- Spennemann, D. H. R. (2026b). *Education governance and datafication. A data set of references hallucinated by generative AI*. School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University. [CrossRef]
- Spennemann, D. H. R. (2026c). How unique are hallucinated citations offered by generative Artificial Intelligence models? *arXiv*, arXiv:2604.16407. [CrossRef]
- Spennemann, D. H. R. (2026d). *One hundred essays on education governance and datafication written by ChatGPT 5.3*. School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University. [CrossRef]
- Spennemann, D. H. R. (2026e). *Ten essays on education governance and datafication written by ChatGPT 5-mini*. School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University. [CrossRef]
- Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13(1), 14045. [CrossRef] [PubMed]
- Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage.
- Williamson, B. (2026). *Tracing the social half-life of a zombie citation*. Available online: <https://codeactsineducation.wordpress.com/2026/01/30/tracing-the-social-half-life-of-a-zombie-citation/> (accessed on 30 January 2026).
- Williamson, B., & Hogan, A. (2020). *Commercialisation and privatisation in/of education in the context of COVID-19*. Education International Research Brief. Education International Research Brussels.
- Williamson, B., & Piattoeva, N. (2019). Objectivity as standardization in data-scientific education policy, technology and governance. *Learning, Media and Technology*, 44(1), 64–76. [CrossRef]
- Williamson, B., & Piattoeva, N. (2020). Objectivity as standardization in data-scientific education policy, technology and governance. In J. Juliane, & B. Andreas (Eds.), *The datafication of education* (pp. 81–93). Routledge.

- Xu, Z., Qiu, Y., Sun, L., Miao, F., Wu, F., Wang, X., Li, X., Lu, H., Zhang, Z., & Hu, Y. (2026). GhostCite: A large-scale analysis of citation validity in the age of large language models. *arXiv*, arXiv:2602.06718. [\[CrossRef\]](#)
- Yaddaf, M. (2025). How to spot fake scientists and stop them from publishing papers. *Nature*, 646, 793. [\[CrossRef\]](#)

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