

Article

ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis

Björn Kindenberg 

Department of Education, Stockholm University, 114 18 Stockholm, Sweden; bjorn.kindenberg@su.se

Abstract: This study investigates alternative approaches for demonstrating historical understanding in elementary school history education, motivated by challenges to educational institutions posed by increased ChatGPT-related plagiarism. Focused on secondary education, an area with scant research, this study, through sociocultural and linguistic methods of analysis, contrasted human-generated historical narratives with those produced by ChatGPT. It was found that ChatGPT's narratives, while stylistically superior, lacked emotional depth, highlighting a key differentiation from human storytelling. However, despite this differentiation, ChatGPT otherwise effectively mimicked typical discourse patterns of historical storytelling, suggesting that narrative-based writing assignments do not significantly reduce the likelihood of ChatGPT-assisted plagiarism. The study concludes by suggesting that rather than focusing on mitigating plagiarism, educational approaches to ChatGPT should seek to channel its potential for historical narratives into assistance with task design, delivery of content, and coaching student writing.

Keywords: history education; historical understanding; artificial intelligence; writing; secondary school; assessment



Citation: Kindenberg, B. ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis. *Educ. Sci.* **2024**, *14*, 530. <https://doi.org/10.3390/educsci14050530>

Academic Editors: Danial Hooshyar, Roger Azevedo and Raija Hämäläinen

Received: 3 April 2024
Revised: 8 May 2024
Accepted: 11 May 2024
Published: 13 May 2024



Copyright: © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This study investigates alternative approaches for demonstrating historical understanding in history education, in the light of ChatGPT's potentially transformative impact on education. The introduction of ChatGPT, a generative artificial intelligence tool, represents a significant milestone in technological advancement. ChatGPT is adept at generating non-preset, human-like textual content in response to various inputs (so-called 'prompts'), for instance, specific instructions and queries [1]. However, the deployment of ChatGPT in educational settings has raised some critical questions regarding its implications. While educators seem hopeful about the prospects of dynamic, AI-enhanced learning environments, they are simultaneously concerned over issues like diminished emphasis on critical thinking and the surge of plagiarism [2]. In current discourse, the risk of plagiarism seems to overshadow potential pedagogical benefits that ChatGPT could bring to learning environments [3]. The tool's evolving proficiency in mimicking academic discourse further exacerbates these concerns. As systems like ChatGPT continue to advance through training on diverse data sources (books, reports, poems, blogs), their capacity for producing increasingly sophisticated human-like texts is noted across diverse settings [4], an observation that necessitates reconsideration of current assessment strategies [5]. Prior studies indicate that ChatGPT-generated texts in established academic genres, like argumentative essays, are rated superior to those composed by students [6]. This signals that challenges such as plagiarism are not transient and that schools need to critically interrogate their use of ChatGPT and existing assessment methods [7–9].

In the present study, storytelling is considered as a candidate for alternative assessment methods in history education. Storytelling represents a profoundly human activity, often productively employed as an instructional strategy [10] that draws on resources such as creativity, imagination, and cultural experiences to establish personalized communication

with the audience [11]. This activity is highly regarded by educators for its capacity to engage students creatively with cultural experiences [12]. Unlike the more formulaic structure of argumentative writing, storytelling encourages a diverse and personal discourse pattern, potentially making it less susceptible to machine replication.

This study investigates the potential of storytelling as a less plagiarism-susceptible alternative to traditional writing formats in lower-secondary education. It compares historical narratives generated by ChatGPT with those written by secondary-school students. This comparison considers both the historical understanding and the textual quality of the narratives. Specifically, the study addresses the following research question: How do ChatGPT-generated historical narratives compare to those written by students in terms of historical understanding and writing quality?

2. Experiences with ChatGPT in Education

ChatGPT can be viewed both as a transformative educational game-changer and as an extension of previous AI applications to improve educational practices. In their pre-ChatGPT systematic review, Chiu and colleagues [13] identified features such as personalized interactions, timely feedback, automated grading, and enhanced student motivation as potential benefits of AI in education. While such potential benefits have been ascribed to ChatGPT as well [14], it is evident from earlier reviews, such as Chiu and colleagues', that ChatGPT marks a paradigmatic shift in the application of AI within educational settings. For example, prior studies often focused on AI in specialized instructional contexts—for instance, using AI-powered motion-capture technology in dance education [15] or machine learning to aid the teaching of glomerulopathies [16]. In contrast, ChatGPT enhances accessibility and broadens application possibilities across educational domains through its integration into widely used software tools (e.g., Microsoft's Office 365) and common search engines [17].

Another limitation of previous AI applications was the requirement for educational environments to conform to the technology's capabilities, rather than adapting the technology to meet educational needs. In their review, Chiu and colleagues note a preference among educators and students for a "more user-friendly and effective system that provides meaningful advice over the mechanical repetition of feedback" [13] (p. 8). ChatGPT contrasts with such specialized AI applications—which require relatively advanced and sometimes expensive equipment—by providing an intuitive dialogue-based interface with general areas for application [18].

Although research on applications of ChatGPT in education is rapidly growing, the current understanding of ChatGPT's impact seems primarily informed by literature reviews. For instance, a recent systematic review [19] primarily surveys SWOT analyses [20], reviews [21], and position papers [22], which can perhaps be seen as an indication of a gap in empirical research concerning ChatGPT's effectiveness. The need for empirical studies is further emphasized by the rapid incorporation of ChatGPT into educational practices and its high acceptance among students [23] (Zhai, 2022).

Recently published studies indicate a broad range of ChatGPT applications in educational settings. Rodriguez [24] has outlined several applications that seem to be increasingly in use among early adopters of AI technology in history education, including the utilization of ChatGPT to suggest classroom debate topics, design educational games, design tests, and devise role-playing-based simulations of historical events. Similarly, Cooper [25] has evaluated its application in STEM and found that it demonstrated a strong potential to support science education by generating educational content and assisting in the creation of teaching materials such as units, rubrics, and quizzes, in addition to simulating pedagogically useful conversational exchanges. Other recent studies have further explored diverse educational uses of ChatGPT, such as supporting problem-based learning in medical education [26], improving feedback on oral presentations [27], and enhancing conceptual understanding in STEM education [28].

Despite ChatGPT's ability to facilitate instructional design, implementation, and assessment, concerns remain among educators. For instance, Davis and Lee [29] have noted that ChatGPT lacks the ability to respond to and integrate students' prior knowledge with evolving learning experiences and sometimes fabricates information. Additionally, its ease of access has raised alarms regarding the potential for increased plagiarism [30]. Due to its proficiency in generating academic texts, ChatGPT is believed to potentially render existing assessment formats, like online exams, obsolete [31]. Specifically, the essay assignment is considered highly susceptible to unauthorized use of ChatGPT [32,33]. The traditional format of historical essays, although widely recognized as "the pinnacle of historical writing" [34] (p. 559), has been critiqued for being strictly conventionalized in educational settings [35] and the now standardized 'five paragraph essay' is arguably highly susceptible to AI replication. While in earlier studies AI technology was highlighted for its potential to automate essay scoring [36], the capacity to mimic historical essay discourse patterns now threatens academic integrity by facilitating the generation of texts. This has led to calls for rethinking assessment methods to make them less susceptible to AI influence [8,9,37]. It has been suggested that since ChatGPT (as of yet) struggles with a nuanced understanding of emotions, educators should prioritize emotional dimensions of learning [38]. This paper explores whether historical narratives can serve as an assessment tool that minimizes unauthorized ChatGPT use by emphasizing personalization and emotional depth in the demonstration of historical understanding.

3. Historical Understanding and Narratives

The study explores the written narrative mediation of historical understanding. Here, historical understanding is defined through Barton and Levstik's [39] conceptualization of historical understanding as students' adoption of *historical stances*:

- *Identification*: When students associate themselves with specific persons, people, or events, including personal connections to the past in some form;
- *Analytic*: To critically examine history, seeking for causes and consequences of events, or actions or decisions taken in the past;
- *Moral response*: Making ethically grounded judgments about the past, such as evaluating, condemning, or admiring persons, decisions, historical outcomes, and so on;
- *Exhibition stance*: Demonstrating, organizing, and presenting historical information.

Stances fall into two primary orientations: a *public orientation*, which focuses on society and its structures, and an *individual orientation*, which emphasizes the role of and experiences of individuals in history. A comprehensive historical understanding emerges from integrating these stances and orientations. For example, understanding European imperialism entails adopting an analytic stance, which, in turn, requires students to recognize how social, economic, and ideological structures are intertwined with the actions and decisions of individuals. Often, collectives that have emerged in the past encompass students (e.g., Europeans, the middle class, capitalist societies). Ideally, students will recognize and negatively evaluate the impact of imperialism (e.g., contemporary racist ideologies). Hence, different orientations of stances interrelate.

Barton and Levstik [39] propose that historical stances are mediated by cultural tools, such as narratives. The present study concentrates on *historical narratives*, defined herein as written stories that organize events within a historical context in chronological order. While the term narrative covers a broad range of texts, Rothery and Stenglin [40] articulate their common social purpose as entertaining an audience "by giving the events spoken or written about a significance within their respective fields" (p. 232). Applied to the field of history, the function of a (written) historical narrative is to captivate the reader with a series of historically contextualized events, facilitating a deeper understanding and connection with the past. To this end, authors may employ a diverse array of techniques ranging from the foundational—such as the use of descriptive adjectives to vividly render settings and characters—to the more complex, including the utilization of metaphors and similes. Gardner [41], among others, has provided a comprehensive summary of these techniques.

4. Materials and Methods

This section describes the data collection and analytical procedures used in the investigation.

4.1. Data Collection

The student-written texts analyzed in this study were collected from a broader case study dataset. Following the completion of the case study, these texts were compared with equivalent texts generated by ChatGPT (version 4), using NVivo software (version 12) for qualitative analysis as described below. The preceding case study observed three eighth-grade classes (a total of 49 students) during a five-week unit on early European colonization of Native American, African, and Asian civilizations (15th to 17th century C.E.). The study was conducted in a school in a linguistically diverse and socioeconomically disadvantaged area outside Stockholm, Sweden. Data collection complied with the ethical research protocol set by the Swedish Research Council [42]. Students and their guardians were informed that participation was voluntary and that their written consent to participate could be withdrawn at any time (A translated and anonymized version of the content form can be accessed at figshare.com: 10.6084/m9.figshare.25773279). At the time of the study's inception, institutional guidelines did not require a now-standardized committee review. Despite the absence of a formal review, the study was conducted with strict adherence to ethical principles relevant at the time, ensured through continuous discussions with senior researchers throughout the study.

During the unit, students were given lectures and reading assignments, took part in group discussions, and engaged with various educational resources (e.g., films and websites). For the final assignment, students could choose from several writing formats, including one corresponding to the above definition of historical narratives. This task requested students to write a story set in the era of colonization, for example: "Pretend that you are a sailor onboard Columbus' expedition and retell your experiences". Students were informed that in these stories they should demonstrate relevant historical knowledge. Most students preferred non-narrative format options (informational reports or argumentative essays), finding the narrative task challenging. Three students finished first-person historical narratives, on different topics (in Table 1 denoted S1, S2, and S3, respectively). These texts were selected as data. The texts were written in Swedish and were translated to English for the present study.

Table 1. Texts examined.

Writing Prompt	Student Text	ChatGPT Text Pre-Feedback	ChatGPT Text Post-Feedback
"Pretend that you are a sailor onboard Columbus' expedition * and tell your experiences".	Text S1 1715 words	Text ChatGPT1a 582 words	Text ChatGPT1b 481 words
"Pretend that you are a sailor onboard Magellan's expedition ** and retell your experiences".	Text S2 1301 words	Text ChatGPT2a 546 words	Text ChatGPT2b 470 words
"Imagine that you are a person living in the Inca Empire when the Spanish Conquistadores arrive and conquer your country ***. Retell your experiences".	Text S3 2262 words	Text ChatGPT3a 485 words	Text ChatGPT3b 418 words

* 1492–1493 C.E.; ** 1519–1522 C.E.; *** 1532–1572 C.E.

At the time the case study was conducted, ChatGPT had not been launched. To compare student-written texts with ChatGPT-generated ones, the researcher provided ChatGPT (paid-for version GPT-4) with the same writing prompts that these students had used. Prompts and ChatGPT responses were in English. These prompts were intentionally kept simple, to simulate how an eighth-grade student would presumably prompt ChatGPT. During instruction, students were given continuous support in their writing by their teacher. To simulate this support, ChatGPT was given feedback on its texts, encouraging more developed answers, which resulted in two ChatGPT versions per writing prompt. As

seen in Table 1, initial versions of ChatGPT-generated texts are denoted with the letter a (e.g., ‘ChatGPT1a’), and versions post-feedback with the letter b (e.g., ‘ChatGPT1b’).

4.2. Analysis

The analysis of texts was conducted in two primary steps, one focusing on content comprehension and the other on writing quality. Initially, the texts were examined using a coding framework designed to assess historical understanding. This framework, inspired by Barton and Levstik’s (2004) above-described conceptualization of *historical stances* (coded as I, A, M, and E), distinguished between identification (I), analytical (A), moral response (M) and exhibition of knowledge (E) stances. Each code was further differentiated by the orientation of each stance—either individual (1) or public (2). For example, Columbus’ 1492 expedition portrayed as motivated by personal ambition or greed would be coded A1, whereas a portrayal identifying it as part of European trade expansion would be coded A2. Disapproval of Pizarro’s cruelty would be coded as M1, reflecting its concern with his individual behavior, while criticism of aggressive Spanish policies would be marked as M2. Table 2 displays the coding scheme and exemplifies the application of each code in this initial step of the analysis.

Table 2. Coding scheme for identifying historical understanding.

Stance	Stance Orientation (Code)	Descriptor	Example
Identification	I1	Personal connections to individual actors are made, either by the author assuming the identity of a character, by vivid description of a character, or by explicitly associating with an actor or group of actors.	“I am the sailor, Rodriguez, embarking on Columbus’ ship, the Santa Maria”.
	I2	Actors are recognized as part of historically situated structures, larger groups, or institutions.	“During the 15th century, Europeans such as Columbus set out on global explorations”.
Analytical	A1	In the text, accounts are found for causes and consequences of individual actors’ actions and behavior; not seen in a wider historical context.	“Atahualpa wanted to appear peaceful and was therefore unarmed, so the Conquistadores could easily capture him”.
	A2	Economic, political, or other underlying and/or long-term causes and consequences are accounted for.	“The ease with which Atahualpa was defeated led the Spanish to conquer and colonize the entire vast Inca Empire”.
Moral response	M1	Opinions about individual actors’ either commendable or morally questionable actions and behavior can be found in the text.	“The cruel Pizarro ruthlessly killed the defenseless Atahualpa”.
	M2	The response is generalized, e.g., to current events.	“The cruel treatment of native Americans has continued to the present day”.
Exhibition of knowledge	E1	The text presents historical facts, often detailed and/or with interest for a specific topic.	“Only 18 men returned from Magellan’s expedition”.
	E2	The text presents historical facts and indicates why this information may be of general interest.	“Even today, the sea that Magellan named is called the Pacific Ocean”.

Subsequently, codes were aggregated into two superordinate categories: *historical understanding, level 1*, reflecting texts predominantly containing codes I1, A1, M1, and E1, and *historical understanding, level 2*, marked by a robust presence of codes I2, A2, M2, and E2. This classification enabled a comprehensive assessment of the historical understanding demonstrated in each text. Additionally, instances of factual inaccuracies and anachronisms were noted. This additional examination was conducted partly to account for ChatGPT’s proneness for ‘hallucinations’ [43], partly as a recognition of the risk for *presentism*, when students interpret the past through the lens of contemporary values, and where personal experiences and opinions risk being tied to the writing of first-hand historical narratives [44].

As argued by Bertram and colleagues, “both *meaning* and linguistic *form* are important dimensions in the evaluation of student answers on open tasks” [45] (p. 21). Consequently, a second step of the analysis of texts assessed their narrative qualities, focusing on how linguistic resources were employed to engage and animate the narratives. Basic descriptions (e.g., verbs, adjectives, and emotive language) were distinguished from more sophisticated

literary devices, including metaphors, similes, hyperboles, or idioms. For example, sophisticated writing in the data featured expressions like “We sailed for an eternity” (a hyperbole), “he led with an iron fist” (idiom), “the excitement was palpable” (metaphor), and “an event as unexpected as the Andean winds” (simile).

This assessment differentiated between basic descriptive elements and more sophisticated stylistic devices, including metaphors, similes, hyperboles, and idioms, categorizing the texts into two levels of narrative quality (level 1 and level 2) based on their complexity, with the first level categorizing predominantly using simpler techniques, and the second level encompassing texts with a robust presence of complex stylistic elements. NVivo was utilized to extract percentages representing the prevalence of different coding elements (historical stance orientations, and basic versus sophisticated devices). The analysis culminated in a cross-comparison of the distribution of codes related to historical understanding and narrative quality, respectively. These distributions were visually represented through pie and bar charts, integrating qualitative analysis with quantitative data representation for a comprehensive understanding of the data. To simplify data visualization, coded files from NVivo were exported to Excel. Subsequently, the graphs generated in Excel were then formatted using PowerPoint, to improve their readability. The resulting visualizations are presented in the following section.

5. Results

This section summarizes the results of the data analysis. Refer to Appendix A for details of the analysis and to Appendix B for contrasting examples of student and ChatGPT writing styles. Table 3 provides a comparative overview of the texts based on the categorization of narrative qualities and historical understanding.

Table 3. Texts’ level of narrative quality compared to level of historical understanding.

	Historical Understanding, Level 1	Historical Understanding, Level 2
Narrative Quality, Level 2	GPT1a GPT2a	GPT2b GPT3b
Narrative Quality, Level 1	S2 S3 GPT3a	S1 GPT1b

Table 3 indicates that most ChatGPT-generated texts were stylistically well-constructed narratives. A quote from text ChatGPT2a can be used to illustrate ChatGPT’s writing style:

The days have been hard. Our diet consists mainly of hardtack, salted meat, and fish. Fresh water is rationed. The vastness of the ocean seems endless, and there’s no land in sight. The winds have been both a blessing and a curse. At times, they propel us forward; other times, they stall our progress (ChatGPT2a).

Table 3 further suggests that prior to receiving feedback, ChatGPT’s historical narratives were on par with those of students in terms of historical understanding. This implies that ChatGPT did not initially emphasize historical analysis but did so when prompted.

5.1. Factual Errors, Anachronisms, and Presentisms

The texts mostly avoided factual errors, anachronisms, and/or presentisms. However, some potential misrepresentations were noted. One example was in text S1, where the destination of Columbus’ voyage was incorrectly stated as Japan instead of India. Text S3 inaccurately depicted the Spanish Conquistadors arriving by boat to Machu Picchu, encountering the Inca Empire ruler Atahualpa, while Cajamarca is the historically accurate site of this encounter (like Machu Picchu, it was a landlocked city). Moreover, this text presented a somewhat ‘sugarcoated’ version of historical events, portraying Atahualpa, by historians believed to be an unpopular despot, as a beloved ruler. In this story, Inca citizens

quickly take up resistance in one-man operations reminiscent of Hollywood action films. Thus, this text's version of events could be considered an example of presentism.

In contrast, ChatGPT texts were free from factual errors or cases of presentism, but not anachronisms. An example in text GPT3b involved the nameless narrator, "a denizen of the Inca Empire", inexplicably referring to events transpiring centuries after the Inca downfall: "The silver of Potosí didn't just decorate European homes; it flowed eastwards to Asia, linking global markets, financing wars, and altering the very course of global history". While historically correct, the idea that an inhabitant of the Inca Empire would make such comments undermines the narrative credibility.

5.2. Comparison of Historical Understanding and Writing Quality

Table 1 shows that some ChatGPT-generated texts demonstrated a particular level of excellence in both crafting a coherent historical narrative and demonstrating in-depth historical understanding (For a detailed breakdown of stances and linguistic features, see Table A1 in Appendix A). However, nuanced differences appeared upon closer examination, as detailed in this section.

Figure 1 contrasts two texts about Magellan's voyage; text S2 (student-authored) and text ChatGPT2a (AI-generated, before feedback). In terms of narrative quality, GPT2a displays use of sophisticated stylistic devices. For example, its opening sentence—"As I boarded one of the five ships in Seville, the air was thick with a mixture of excitement and anxiety"—and similes like "The crew is suffering from scurvy, and despair looms over the ships like a dark cloud", demonstrate this.

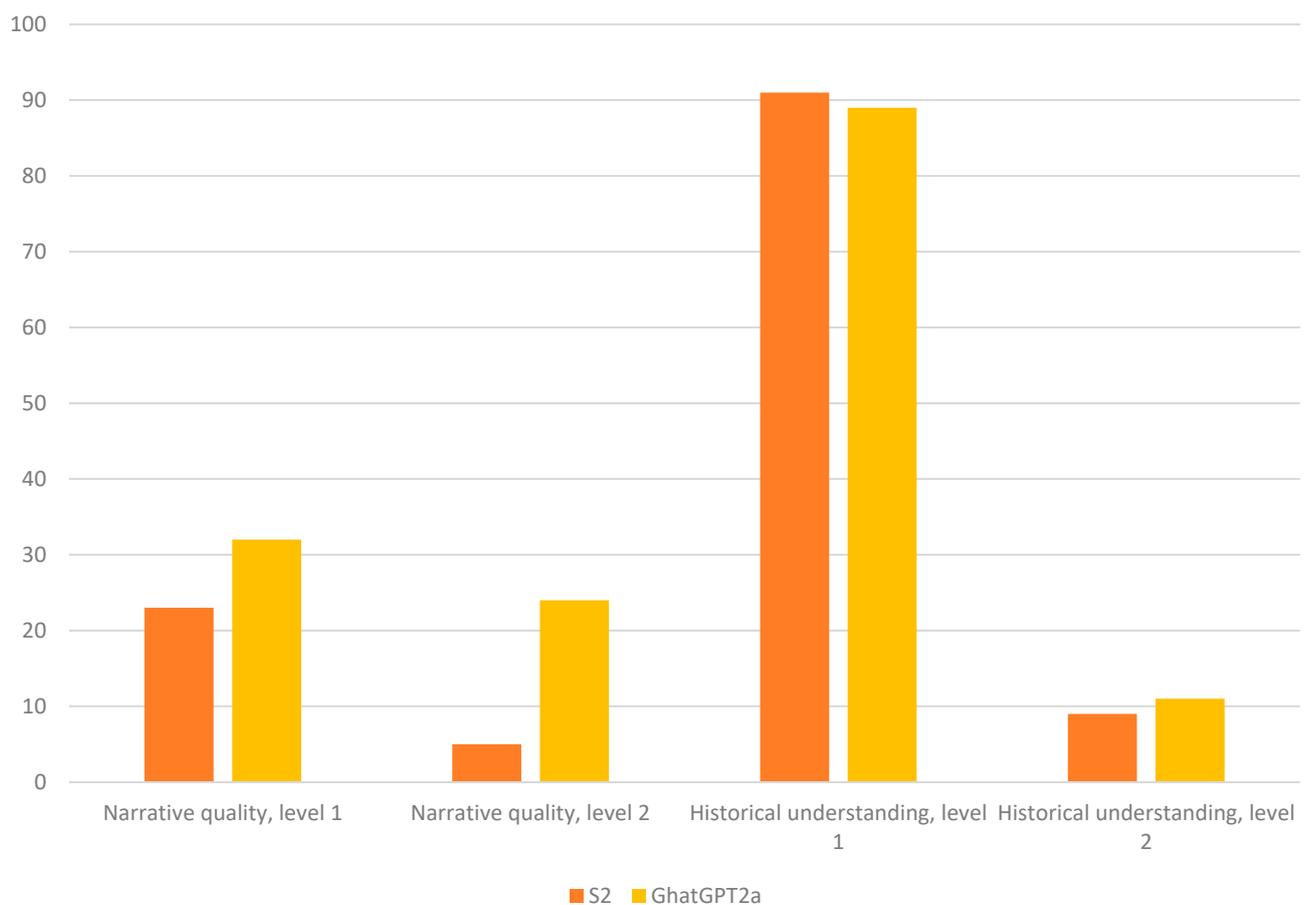


Figure 1. Student-written and AI-generated text.

The student-written text was, likewise, a competently crafted story, commencing *in medias res* with the narrator reading a sign posted by explorer Magellan:

Ferdinand Magellan, born in Sabrosa Portugal is to lead the first ever circumvention of the world. But we're looking for a crew, sign here to take part of this expedition and join this voyage of discovery.

In its use of basic stylistic features, such as adjectives or evaluative language, text S2 resembled text GPT2a, albeit lacking in metaphors and similes. In terms of historical understanding both texts were on par. In each text, an individually oriented identification stance (I1) was prominently displayed, which is reasonable given the writing task. The exhibition stance, with an individual orientation (E1), was more prominent in S2 (42%) than in the much shorter GPT2a (33%), indicating richer historical detail in the student text.

Figure 2 demonstrates an even more pronounced contrast between student text and pre-feedback ChatGPT text. The GPT1a text, rich in historical facts like dates and names, primarily displayed individual orientations of the exhibition and identification stances towards historical events. For example, the reasons given for Columbus' voyage were his personal desire "to find a western route to Asia", "promises of riches", and "the thrill of the discovery" as opposed to more underlying and structural causes for European colonialism. While text GPT1a mentioned the wider significance of the event, commenting that "the horizon has expanded and with it the known world", its focus was on individual actors' motives, such as the allure of discovery.

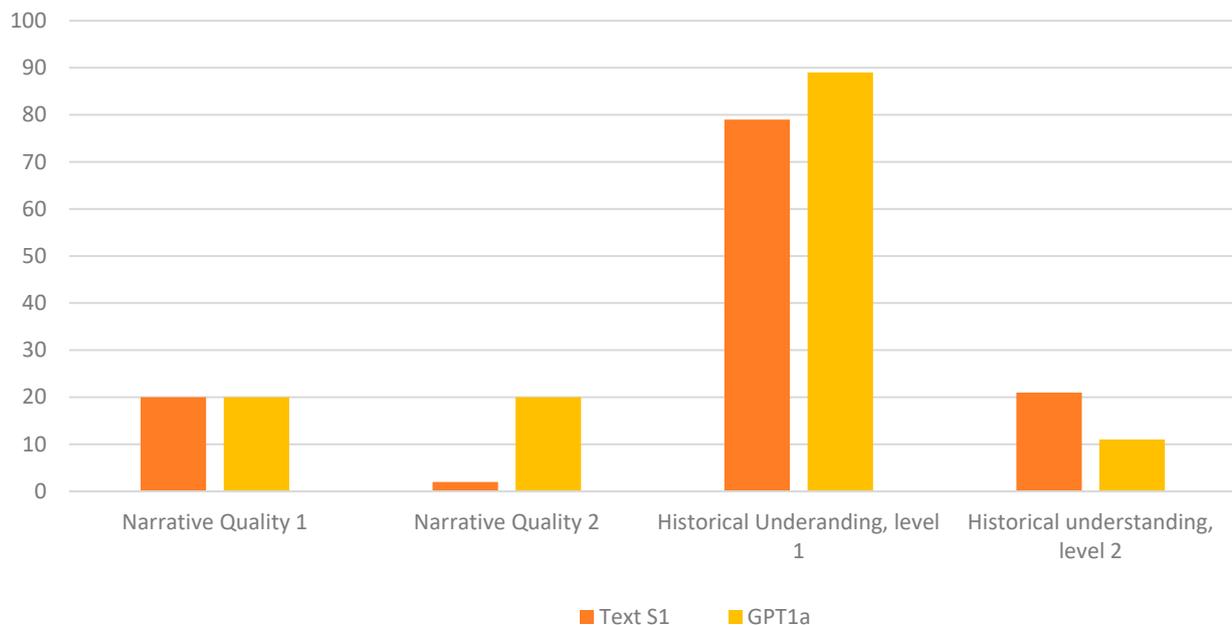


Figure 2. Student-written texts and corresponding ChatGPT-generated texts, compared.

In contrast, student-written text S1 offered a more complex analysis that extended beyond individual experiences to encompass broader European economic and political ambitions. This was enabled by the student's use of retrospective narration, a technique that, interestingly, was never used by ChatGPT. Reflecting thirty years later, narrator Sebastian could credibly comment on the expedition's long-term impacts, including Columbus' role in future European expansion:

The purpose of these trips was to find something we didn't know before. He [Columbus] sought money from Portugal and Spain. Spain gave Columbus support (money, boat, etc.) and he was then able to conquer riches but eventually conquer lands in Spain's name. My name is Sebastian Rizzo and I have been on this strange journey.

Here, the author makes critical observations about the implications of the voyage that Sebastian has been on, including the eventual conquering of lands "in Spain's name".

This identification of a larger collective and their historical impact reflects the adoption of historical stances A2 and I2 (analysis and wider identification), which the retrospective narrative enabled.

Conversely, text GPT1a (titled “Matteo’s journal”) was a ‘real-time’ travel diary, as exemplified here:

3 August 1492

Today, we departed from the port of Palos in Spain, embarking on a voyage that many called mad. I joined Admiral Cristóbal Colón, better known as Christopher Columbus, on this journey to find a western route to Asia.

While the travel diary format allowed ‘Matteo’ to make observations about economic motives for colonialism—reflected in the excerpt as ‘a western route to Asia’—these observations did not integrate as seamlessly with the narrative as was done in text S1.

The difference between these texts was further accentuated by text S1’s appended section with personal commentary. This section not only reflected on the journey’s long-term contribution to the development of Euro-centric mindsets but also effectively situated Columbus within a broader European context and condemned present-day racist attitudes as a legacy of European expansion (analyzed as an adoption of stances A1, I2, and M2).

This section was not required by the teacher, but the student seemed to have experienced a need to delve deeper into the implications of this historical event. The feedbacked ChatGPT text (GPT1b, curiously titled “The analytical journal of Matteo”) employed a different approach to incorporate historical analysis:

Today marks a pivotal juncture in history as we departed from Palos. Amid the European race for spices, silks, and other riches of the East, our voyage represents not just a personal quest but a broader economic ambition of Spain. Our nation seeks alternatives to the treacherous and long-established Silk Road, dominated by the Ottomans. Many of my shipmates, enticed by the potential of wealth and new trade routes, voice their concerns about the unknown, reflecting the larger societal fear of the uncharted.

The excerpt shows ChatGPT’s extensive use of denoting verbs like ‘mark’, ‘represent’, and ‘reflect’, at the expense of words expressing thoughts and feelings (e.g., ‘exclaim’, ‘fear’, ‘think’). The excerpt reflects the pattern that ChatGPT texts revised for deeper analysis became less engaging as narratives in the sense that dispassionate observations about emblematic events took precedence over vivid storytelling. This pattern is evident in Figure 3, where different versions of texts about the fall of the Inca empire are compared.

The first and second bar reflect how ChatGPT in its iteration replaced basic descriptors like adjectives and verbs, which typically convey thoughts, emotions, and experiences, with more advanced stylistic techniques. When deepened analysis was requested (bars three and four), the revised ChatGPT demonstrated an increased use not only of sophisticated stylistic devices but also of historical stances with public orientations. However, this revision resulted in a text that was noticeably less personal in tone.

The shift from engaging narratives to analytically detached texts is evident when comparing three excerpts from stories about the arrival of the Spanish Conquistadors to the Inca society. The first sample, from student text S3, vividly describes the moment:

Like every day, me and my family were out on the terraces tilling the land. Suddenly we saw the Inca and his army heading towards the beach. People followed and I left the shovel I was holding and looked for my friend. I found her and we walked towards the beach. There we saw boats and a large group of men jumping out of them. The leader of the strangers began to introduce himself: ‘My name is Pizarro. We come from Spain’.

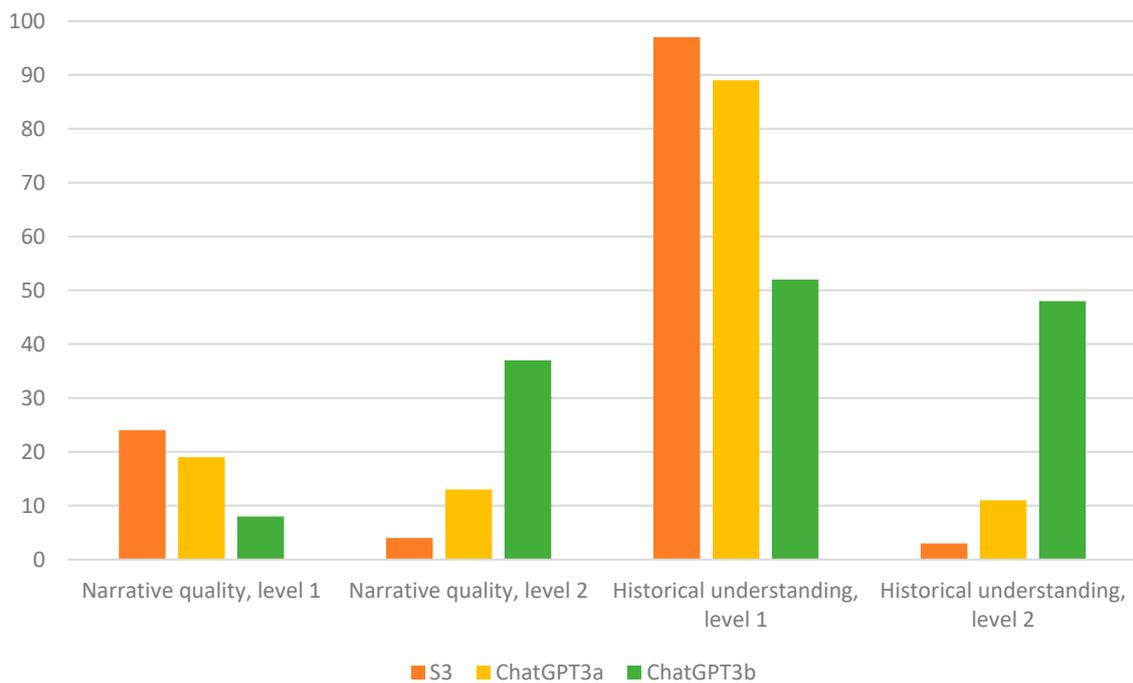


Figure 3. Student-written texts and corresponding ChatGPT-generated texts, compared.

This student’s text, notably longer than the AI-generated versions (Table 1), was rich in historical detail and personal experience, adopting both exhibition and identification stances, and individually oriented (E1, I1). In contrast, the ChatGPT text (GPT3a) offered more complex writing, using idiomatic expressions:

When they came to our village, I saw them with my own eyes: metal-clad men with long beards and avaricious eyes, led by one named Francisco Pizarro. Their language sounded alien, but some acted as interpreters. With their superior weaponry, and often by employing deceit, they began to subdue village after village.

GPT3a matched S3 in writing quality and historical insight (e.g., noting the significance of superior weaponry), incorporating both factual elements and personal reflections (e.g., “avaricious eyes”). However, when prompted for deeper historical analysis, the revised ChatGPT text (GPT3b) became markedly impersonal:

Francisco Pizarro and his cohorts, upon setting foot in our lands, were not just mesmerized by our golden artifacts but also recognized the economic potential of this civilization. The conquest was not just a quest for riches, but a strategic move in the larger geopolitical chessboard, enabling Spain to enhance its stature in the European power dynamics.

Here, the use of the word “our” suggests an identification with the Inca narrator but it is unclear whether this pronoun refers to a specific group or to a broader Inca society. While GPT3b’s observation about the ‘encomienda’ system is eloquent, it conveys little emotional depth and the text seem distanced from—rather than empathizing with—the lived experiences of individuals and collectives in the past.

Although there were variances between student-written texts, a notable distinction was observed in text S3. Unlike texts S1 and S2, who prioritized historical analysis, text S3 focused on historical perspective-taking and empathy, as illustrated by the high proportion of moral and identification stances in this text, as shown in Figure 4. In this figure, the inner circle segment indicates historical stances, which are then broken down into level 1 and 2 in the outer circle segment.

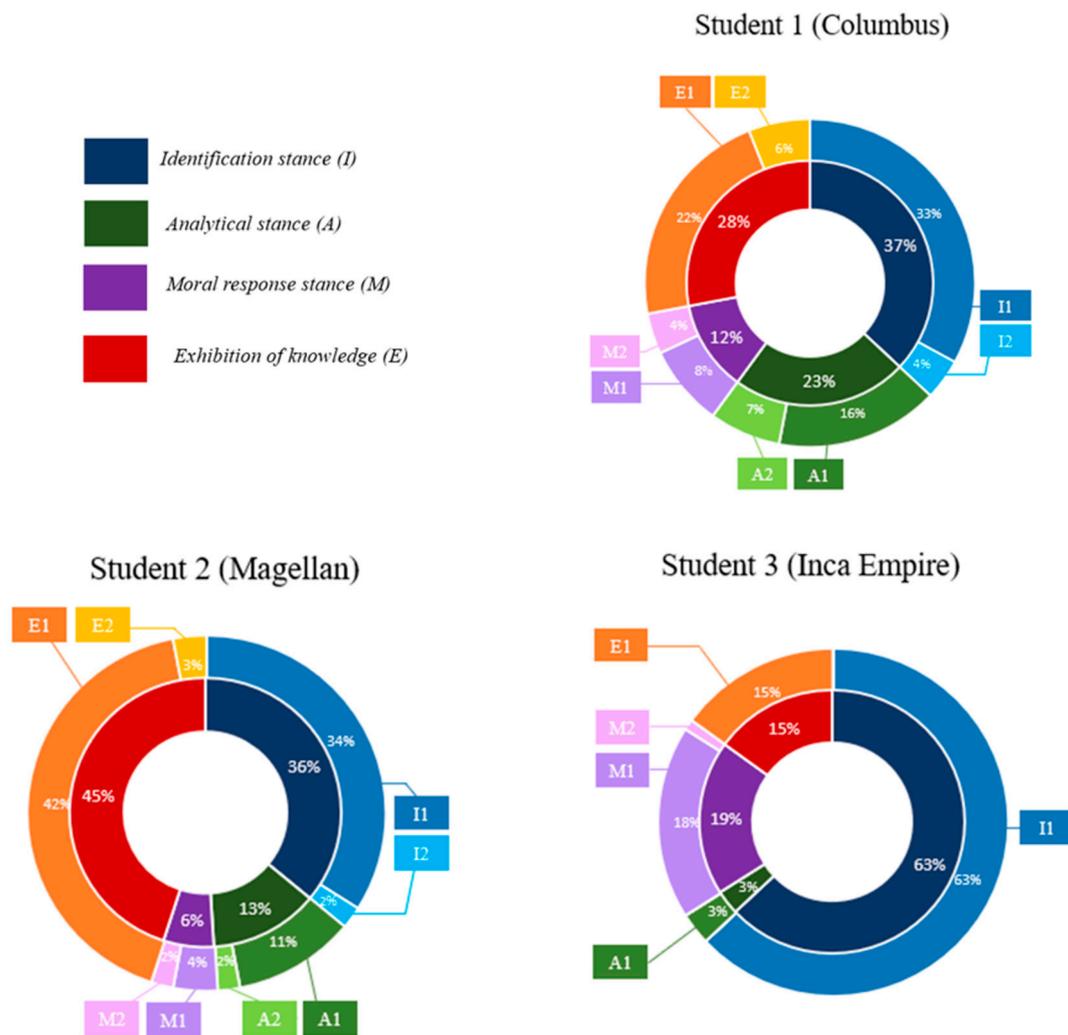


Figure 4. Historical stances in student-written texts.

In text S3, an authorial identification with the young female Inca narrator, paired with a pronounced disdain for the invading Conquistadors, conveyed a strong sense of moral outrage. A quote from the text, “I feel a lump in my throat. It feels as if someone is choking me”, highlights how the author amplified her narrator’s emotional turmoil in response to vivid depictions of the Conquistadors’ brutality (in this example, the narrator’s response was prompted by the Conquistadors’ maiming of her best friend). Such pronounced emotive involvement was not evident in texts S2 and S1, where the authors instead used their respective narrator’s personal experiences rather to interpret and dissect broader historical contexts. This is exemplified effectively in an excerpt from S1:

The large mirror captures my bright green eyes shining. My dirty hands that I hide under my cap. My black hair camouflages into the wall behind me. This is more than what I asked for. We Spaniards didn’t have much against Magellan who was Portuguese. We have been competing with the Portuguese for many years and will continue to do so for eternity. How the king can have nothing against him, I do not know.

In this passage, the narrator’s considerations about his modest beginnings contrast with Magellan’s stature. These thoughts transition into a broader contemplation on the trade rivalry between Spain and Portugal symbolized by the characters. The occurrence of such instances—where personal experiences leverage broader historical analysis—is

evident in the robust presence of analytical stance adoptions, depicted in Figure 4. Moreover, the chart indicates that text S2, relative to other texts, incorporated a substantial degree of historical facts (the exhibition stance). The factual presentation blended with the narrative. For example, as the narrator in text 2 recounted his experience boarding Columbus’ ship, the Santa Maria, he also commented on the shipbuilding techniques of that era.

Similar to student-written texts, ChatGPT-generated ones exhibited variance in displayed historical understanding, as illustrated in Figure 5 (which shows ChatGPT texts prior to receiving feedback).

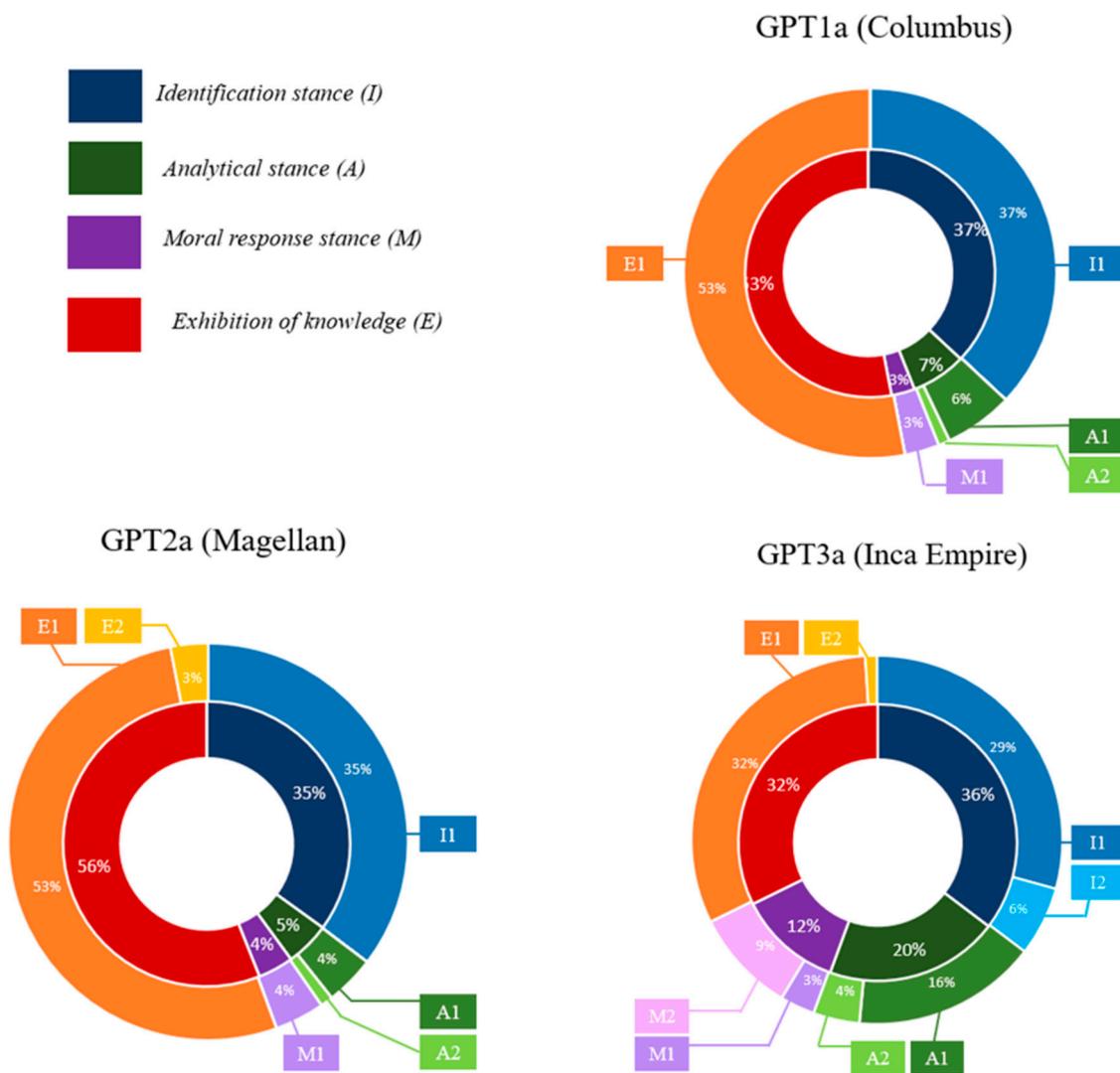


Figure 5. Historical stances in ChatGPT-generated texts (pre-feedback).

All texts adopted a personalized perspective—encouraged by the first-person writing assignment—but differed in emphasis on historical analysis. Notably, the narrative in text ChatGPT3a, which dealt with the Spanish invasion of the Inca Empire, was marked by an approach where historical analysis was significantly emphasized, as shown in the following excerpt:

The Spaniards brought not only their violence but diseases unknown to our people. Many of my kin succumbed to smallpox and other foreign ailments. The great empire which took generations to build was disintegrating before our eyes.

Although the excerpt signals disapproval of the violent demise of an empire that “took generations to build”, it maintains a detached tone. When prompted for deeper analysis of the events, the revised version (ChatGPT3b) incorporated comments such as the following:

The forced imposition of Christianity wasn’t just about religious superiority. It aimed to dismantle our societal structures and beliefs, making us malleable to their rule and worldview.

While moral critique is voiced in this excerpt, its abstracted tone lacks emotional depth. This lack—at times even absence—was a significant characteristic of ChatGPT-generated texts (see Figure 5 and Appendix A) and a marked distinction between human- and machine-generated narratives. This distinction is further explored in the ensuing Discussion Section, which considers the broader implications of these findings.

6. Discussion

The present study undertook a comparative analysis of the proficiency in historical narrative creation between ChatGPT and lower-secondary students, grounded in the assumption that storytelling, inherently reliant on uniquely human capabilities such as imagination and empathy, could be less susceptible to plagiarism in the form of artificial story creation. A preliminary conclusion from the comparison is that storytelling is likely to offer only limited deterrence against plagiarism since personalized historical narratives were effectively simulated by ChatGPT. Nevertheless, the analysis revealed nuanced differences that merit evaluation. One of the key findings from the analysis is that ChatGPT’s creation of historical narratives was not devoid of challenges. Although ChatGPT-generated texts were superior in writing quality, ChatGPT’s initial historical narratives (before feedback) did not immediately surpass students’ demonstrated understanding. However, when prompted, ChatGPT demonstrated deeper understanding. This might align with previous observations that while history educators commend the clarity and coherence of ChatGPT’s historical arguments, they find its exposition of factual knowledge occasionally lacking in depth [46]. This, in turn, might be explained by the constraints, documented in history education research [44,47], that personalized historical narratives inflict on historical interpretations. While the temptation for students to utilize ChatGPT for unsolicited writing assistance is likely, it remains uncertain whether they would accurately assess the quality of such historical narratives.

Another issue identified in ChatGPT-generated texts, as well as in student narratives, was the occurrence of presentism (anachronistic interpretations of historical events), anachronisms, and factual errors, a long-discussed risk in history education [48,49]. These instances were, however, not paradigmatic examples of misunderstanding but rather indicators of what has been labeled *narrative truth* [50], the notion that occasional historical inaccuracies within a narrative do not necessarily diminish its overall historical interpretation. This notion was evident in a student’s portrayal of Columbus’s expedition as destined for Japan, a statement which is not a clear instance of factual error since Columbus’s intentions did include exploring sea routes to Japan (and not only to India as is sometimes understood). Another student text displayed clear misconceptions about the geographical extent of the Inca Empire, in addition to other emotionally charged but probably historically inaccurate details. However, although these additions were possibly influenced by contemporary media consumption and a wish for ‘melodrama’, it did effectively convey the brutality of the Spanish conquest. Notably, while ChatGPT-generated texts navigated similar historical events, they exhibited a lower prevalence of factual misrepresentations.

A pronounced difference between student and ChatGPT authors was that the latter struggled with emotional engagement in narratives. In fact, when prompted to add more layers of moral evaluation, ChatGPT responded that this was beyond its algorithms. The difference can be considered in the light of Tirado-Olivares and colleagues’ study [6], who found that educators’ assessments of ChatGPT- versus student-written historical essays favor dispassionate machine-generated texts due to their coherence and analytical depth. Participants in Tirado-Olivares and colleagues’ study largely agreed that students were

better at conveying emotions but that this was not to their advantage in terms of assessment, since emotional detachment is more beneficial for historical arguments [51]. However, if emotive and ethical stances towards past events is considered a legitimate curricular goal, storytelling's emphasis on emotions and empathy align with educational goals [52,53]. In other words, provided that educators recognize the potential of emotional engagement with the past, as theorized by Barton and Levstik [39], ChatGPT's reluctance to incorporate moral perspectives may reduce its appeal as a tool for plagiarism.

However, despite these challenges associated with ChatGPT, the employment of historical fiction writing in traditional school history will likely not deter students from unauthorized AI-powered assistance. In the study, ChatGPT-generated texts demonstrated levels of historical understanding equal to or exceeding that of students. Unless moral dimensions of historical understanding are given prominence, the high acceptance of ChatGPT among students [11], combined with their readiness to utilize it for diverse writing tasks [47], suggest that the implementation of historical fiction assignments alone might not suffice to prevent the use of ChatGPT in schools. Nevertheless, historical fiction emerges as a valuable instructional resource, with ChatGPT enabling promising venues for enhanced historical understanding. For instance, ChatGPT's ability to simulate eyewitness accounts of significant events, as exemplified in the present study, may help engage students with the past by capitalizing on 'the didactic function of narratives' [54]. This possibility aligns with the growing emphasis on empathetic and engaged learning in history education, where the goal is not just the acquisition of historical facts but also the development of a deeper, more nuanced understanding of historical events and their human impact.

Moreover, the potential of ChatGPT as a tool for coaching historical writing is notable. It can provide templates, writing assistance, and personalized feedback [24,55], aspects that educators find valuable [50]. Tools powered by generative artificial intelligence have been noted as remarkably proficient in assessing argumentative qualities in students' historical essays [47], suggesting unexplored territory for investigations into how these tools could potentially enhance historical narratives for educational purposes. By balancing the capabilities of AI with the insights and guidance of educators, there is an opportunity to not only mitigate the challenges posed by unauthorized AI use but also to leverage these technologies to deepen students' historical understanding and writing proficiency.

In summary, utilizing historical writing to assess students' knowledge of the past is unlikely to prevent AI-enhanced plagiarism. In educational contexts driven by accountability demands, complex human phenomena such as 'history' and 'narratives' are frequently reduced to segments of knowledge that are teachable primarily through written discourse. As a result, AI models like ChatGPT are apt to become proficient in replicating educationally expected discourse patterns, including personalized narratives, and issues of plagiarism will likely persist. To address such issues, recent scholarly debates have called for an epistemological re-evaluation of plagiarism in the era of ChatGPT. Drawing on educational philosopher Paolo Freire's work, Siblin [56] suggests that educators should acknowledge their 'practical ignorance' about the lived experiences of their students and seek to orient their teaching and assessment towards these experiences, rather than towards standardized assessment criteria. McIntire, Calvert, and Ashcraft [57], citing pragmatist philosopher John Dewey, argue that plagiarism is untenable from a pragmatist perspective, as it defeats the purpose of learning—an issue teachers should emphasize to their students. Such emphases on the higher goals of education, rather than emphasis on control mechanisms and plagiarism, tie into Gert Biesta's [58] notion of *subjectification*—the process by which individuals come to understand themselves as autonomous agents (see [58] for an extended discussion on ChatGPT in education, related to subjectification). A worthwhile direction for future research would be exploring how such epistemological and philosophical notions can be translated to instructional practices in educational settings. If embraced rather than opposed, machine tools like ChatGPT can enrich student engagement and, paradoxically, foster a more humanistic understanding of history. As stated by ChatGPT itself, in one of the numerous human-to-machine interactions found online: "AI can be a powerful tool to

augment human abilities, but it should not be seen as a replacement for human thought, judgment, and decision-making” [59]. The present study underscores the need for further explorations of how this dynamic can be negotiated in instructional practice.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study since, at the time of the study’s inception, institutional guidelines at Stockholm University did not require a now-standardized committee review. Despite the absence of a formal review, the study was conducted with strict adherence to ethical principles relevant at the time, ensured through continuous discussions with senior re-searchers throughout the study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data for this study can be made available upon request. Please contact the author.

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

Appendix A. Details of Text Analyses

The calculation methods for percentages vary between narrative quality and historical understanding. For narrative quality, the coding process focused on specific elements within the texts—individual words and expressions, like certain adjectives and idiomatic phrases. This selective coding is the reason why the percentages for level 1 and level 2 narrative qualities do not sum to 100%. Conversely, in coding for historical understanding, all text components were included, which explains why the percentages for different levels of historical understanding collectively equal 100%.

Table A1. Coding of the data, in percentages.

Text	Narrative Techniques			Historical Understanding						
	Basic	Advanced	I1	I2	A1	A2	M1	M2	E1	E2
GPT1a	20	20	37	0	6	1	3	0	53	0
GPT1b	11	21	30	5	9	27	0	0	27	2
GPT2a	32	24	35	0	4	1	4	0	52	3
GPT2b	15	30	26	13	17	21	0	2	19	2
GPT3a	19	13	29	6	16	4	3	9	31	1
GPT3b	8	37	17	5	14	29	2	12	17	5
Student 1	20	2	33	4	16	7	8	4	22	6
Student 2	23	5	34	2	11	2	4	2	42	3
Student 3	24	4	63	0	3	0	18	1	15	0

Appendix B. Two Sample Texts: Student and ChatGPT

Student text S3 (below is an extract from the text) was translated from Swedish to English. ChatGPT-generated text GPT3a is shown below as originally generated in English.

Extract 1, from Text S3:

“I can’t let go”

I’m Pagie and I live with my brother Toby, mum and dad on the outskirts of town in Machu Picchu. We live in a simple stone house with two rooms. We are farmers and usually grow corn, avocados and chilies.

We can neither understand nor interpret “Quipu”, our messaging system. Only those who can afford schooling and have access to schools can interpret it. But going to school requires that a lot of services and barter be made, and we must support ourselves first and

foremost. I wish I could go to school. But I have to content myself with growing corn. And with standing on a staircase and enjoy the beautiful view of the city.

Besides growing on the terraces, we grow on the priests' land, in service to them. My family belongs to the "Inca people", our Inca is called "Atahualpa". We respect him, because he is the son of the sun god, and we listen to them. The city has many buildings: temples, fortresses, wide roads and aqueducts.

Guests arriving

Like every day, me and my family were out on the terraces tilling the land. Suddenly we saw the Inca and his army heading towards the beach. People followed and I left the shovel I was holding and looked for my friend. I found her and we walked towards the beach. There we saw boats and a large group of men jumping out of them. The leader of the strangers began to introduce himself:

"My name is Pizarro. We come from Spain".

The Inca began to introduce himself and you could tell that everyone was curious. We were all interested in who the Spaniards were. Me and my friend Happy were very excited, but also a little scared. The Spanish were armed with gunpowder and weapons, and we had no guns.

"What if they come and hurt us", I said to Happy and she replied: "No, they won't. They are our friends now".

Atahualpa invited them to his palace, they were given food and water. Then I did feel that they are our friends.

The morning after our guests had arrived, I heard a scream, a scream of fear and pain. I ran towards where the scream came from. I arrived at Happy's house. I run in and there she is sitting on the floor. I see a lot of blood on the floor, it's a big pool and I see it's coming from her left foot. Then I realize she her toes have been cut off. I run up to her and ask her what happened.

"The new ones", she replies.

I don't quite understand what she means and then she says: "It was two Spaniards who did it". I sit next to her and ask how she is doing. Although I already know the answer.

"I can't understand that some of our guests, our friends, have done it" I answer, although they are not our friends now, I think.

I feel a lump in my throat. It feels as if someone is choking me.

Extract 2, Text GPT3a:

In the heart of the Andes, under the embrace of Inti, the Sun God, my life took a drastic turn that was as unexpected as the Andean winds.

Before the arrival of the foreigners, I lived in a thriving village near Cusco, the capital of our great Inca Empire. Our ruler, the Sapa Inca, was revered as the living embodiment of Inti on Earth. Life was harmonious; we worked the terraced farms, prayed to our deities, and celebrated our festivals with music, dance, and feasts.

One day, whispers reached our village about the mysterious "Viracochas" from the East. They had strange skin, rode fierce four-legged beasts, and had thundering weapons that spat fire and death. By the time they reached Cusco, many had already bowed to their power or had been brutally silenced.

When they came to our village, I saw them with my own eyes: metal-clad men with long beards and avaricious eyes, led by one named Francisco Pizarro. Their language sounded alien, but some acted as interpreters. With their superior weaponry, and often by employing deceit, they began to subdue village after village.

Cusco fell. Our beloved Sapa Inca, Atahualpa, was captured under treacherous circumstances. I witnessed the sorrow of our people as we were told he was executed, even after a room full of gold and silver was offered for his release.

The invaders imposed their religion upon us, dismissing our deities and sacred practices. They destroyed our temples and erected crosses in their place. Many of our priests were persecuted, and the Sun Virgins of the Coricancha were forced to abandon their sanctity.

The Spaniards brought not only their violence but diseases unknown to our people. Many of my kin succumbed to smallpox and other foreign ailments. The great empire which took generations to build was disintegrating before our eyes.

Resistance was scattered but persistent. Some villages fled to hidden parts of the mountains. Others took arms, leading rebellions against the intruders. But with their horses, guns, and foreign allies, the Spaniards seemed invincible.

I, along with many others, was coerced into a system called “encomienda”. We were forced to labor in mines, fields, or building projects for the Spanish. The mines, especially the silver ones like Potosí, were a death sentence. The air was thick with dust, and the dark tunnels were a stark contrast to the open fields I once tilled.

Despite the despair, hope simmered within our hearts. Stories of secret gatherings, where the old rituals were practiced and plans for resistance were hatched, circulated like precious treasures. The memory of our glorious past, and the dreams of a liberated future, fueled our will to survive.

Even today, as I behold the changed landscape, I know that beneath the Spanish churches, our temples still stand. Within our hearts, Inti’s light remains undiminished. And as long as there’s a single Incan soul alive, our legacy will never truly perish.

References

- Hsu, Y.C.; Ching, Y.H. Generative artificial intelligence in education, part one: The dynamic frontier. *TechTrends* **2023**, *67*, 603–607. [CrossRef]
- Mohamed, A.M. Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL Faculty Members. *Educ. Inf. Technol.* **2024**, *29*, 3195–3217. [CrossRef]
- Kasneji, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günemann, S.; Hüllermeier, E. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
- Visual Oak. 7 Best AI Copywriting Software Tools—W/Free Options 2023. Visual Oak Blog. Available online: <https://www.visualoak.com/best-ai-copywriting-software/> (accessed on 1 March 2024).
- Herbold, S.; Hautli-Janisz, A.; Heuer, U.; Kikteva, Z.; Trautsch, A. AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. *arXiv* **2023**, arXiv:2304.14276.
- Tirado-Olivares, S.; Navío-Inglés, M.; O’Connor-Jiménez, P.; Cózar-Gutiérrez, R. From Human to Machine: Investigating the Effectiveness of the Conversational AI ChatGPT in Historical Thinking. *Educ. Sci.* **2023**, *13*, 8. [CrossRef]
- Trust, T.; Whalen, J.; Mouza, C. ChatGPT: Challenges, Opportunities, and Implications for Teacher Education. *Contemp. Issues Technol. Teach. Educ.* **2023**, *23*, 1–23.
- Ma, G. Chance or Challenge: The Role of ChatGPT in History Teaching and Historical Research in Higher Education. In Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023), Qingdao, China, 21–23 July 2023; Atlantis Press: Amsterdam, The Netherlands, 2023; pp. 869–874.
- Nguyen, Q.H. AI and Plagiarism: Opinion from Teachers, Administrators and Policymakers. In Proceedings of the AsiaCALL International Conference, Da Nang, Vietnam, 25 November 2023; Volume 4, pp. 75–85.
- Mistry, A. The art of storytelling: Cognition and action through stories. *Int. J. Arts Sci.* **2017**, *9*, 301–324.
- Yilmaz, R.M.; Goktas, Y. Using augmented reality technology in storytelling activities: Examining elementary students’ narrative skill and creativity. *Virtual Real.* **2017**, *21*, 75–89. [CrossRef]
- Hawkey, K. Narrative in classroom history. *Curric. J.* **2004**, *15*, 35–44. [CrossRef]
- Chiu, T.K.; Xia, Q.; Zhou, X.; Chai, C.S.; Cheng, M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100118. [CrossRef]
- Sidiropoulos, D.; Anagnostopoulos, C.N. Applications, challenges and ethical issues of AI and ChatGPT in education. *arXiv* **2024**, arXiv:2402.07907.
- Wang, Y.; Zheng, G. Application of artificial intelligence in college dance teaching and its performance analysis. *Int. J. Emerg. Technol. Learn. (IJET)* **2020**, *15*, 178–190. [CrossRef]

16. Aldeman, N.L.S.; de Sá Urtiga Aita, K.M.; Machado, V.P.; da Mata Sousa, L.C.D.; Coelho, A.G.B.; da Silva, A.S.; da Silva Mendes, A.P.; de Oliveira Neres, F.J.; do Monte, S.J.H. Smartpathk: A platform for teaching glomerulopathies using machine learning. *BMC Med. Educ.* **2021**, *21*, 1–8. [[CrossRef](#)] [[PubMed](#)]
17. Rudolph, J.; Tan, S.; Tan, S. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *J. Appl. Learn. Teach.* **2023**, *6*, 364–389.
18. Lim, W.M.; Gunasekara, A.; Pallant, J.L.; Pallant, J.I.; Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manag. Educ.* **2023**, *21*, 100790. [[CrossRef](#)]
19. Patrício, M.R.; Gonçalves, B.F. ChatGPT: Systematic Review of Potentials and Limitations in Education. In Proceedings of the International Conference on Information Technology & Systems, Temuco, Chile, 24–26 January 2024; Springer Nature: Cham, Switzerland, 2024; pp. 339–348.
20. Farrokhnia, M.; Banihashem, S.K.; Noroozi, O.; Wals, A. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innov. Educ. Teach. Int.* **2023**, *64*, 1–15. [[CrossRef](#)]
21. Rasul, T.; Nair, S.; Kalendra, D.; Robin, M.; de Oliveira Santini, F.; Ladeira, W.J.; Sun, M.; Day, I.; Rather, R.A.; Heathcote, L. The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *J. Appl. Learn. Teach.* **2023**, *6*, 1–15.
22. Santandreu-Calonge, D.; Medina-Aguerrebera, P.; Hultberg, P.; Shah, M.A. Can ChatGPT improve communication in hospitals? *Prof. De La Inf./Inf. Prof.* **2023**, *32*, 1–16. [[CrossRef](#)]
23. Zhai, X. ChatGPT user experience: Implications for education. 2022. SSRN [Formerly, Social Science Research Network]. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418 (accessed on 30 March 2024).
24. Rodríguez, A.C. Reinventing the Teaching of Early Modern History in Secondary School: The use of ChatGPT to Enhance Learning and Educational Innovation. *Stud. Hist. Hist. Mod.* **2023**, *45*, 101–145. [[CrossRef](#)]
25. Cooper, G. Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *J. Sci. Educ. Technol.* **2023**, *32*, 444–452. [[CrossRef](#)]
26. Divito, C.B.; Katchikian, B.M.; Gruenwald, J.E.; Burgoon, J.M. The tools of the future are the challenges of today: The use of ChatGPT in problem-based learning medical education. *Med. Teach.* **2024**, *46*, 320–322. [[CrossRef](#)] [[PubMed](#)]
27. Kostka, I.; Toncelli, R. Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *TESL-EJ* **2023**, *27*, 1–19. [[CrossRef](#)]
28. Vasconcelos, M.A.R.; Santos, R.P.D. Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *arXiv* **2023**, arXiv:2305.02202. [[CrossRef](#)] [[PubMed](#)]
29. Davis, R.O.; Lee, Y.J. Prompt: ChatGPT, Create My Course, Please! *Educ. Sci.* **2023**, *14*, 24. [[CrossRef](#)]
30. Yang, H. How I use ChatGPT responsibly in my teaching 2023. *Nature*. Available online: <https://www.nature.com/articles/d41586-023-01026-9> (accessed on 30 March 2024).
31. Susnjak, T. ChatGPT: The end of online exam integrity? *arXiv* **2022**, arXiv:2212.09292.
32. Graham, F. Daily briefing: Will ChatGPT kill the essay assignment? *Nature* **2022**. [[CrossRef](#)] [[PubMed](#)]
33. Kalota, F. A Primer on Generative Artificial Intelligence. *Educ. Sci.* **2024**, *14*, 172. [[CrossRef](#)]
34. Nokes, J.D.; De La Paz, S. Writing and argumentation in history education. In *The Wiley International Handbook of History Teaching and Learning*; Metzger, S.A., Harris, L.M., Eds.; Wiley-Blackwell: Hoboken, NJ, USA, 2018; pp. 551–578.
35. Warner, J. *Why They Can't Write: Killing the Five-Paragraph Essay and other Necessities*; JHU Press: Baltimore, MD, USA, 2018.
36. Kumar, V.; Boulanger, D. Explainable automated essay scoring: Deep learning really has pedagogical value. *Front. Educ.* **2020**, *5*, 572367. [[CrossRef](#)]
37. Gill, S.S.; Xu, M.; Patros, P.; Wu, H.; Kaur, R.; Kaur, K.; Fuller, S.; Singh, M.; Arora, P.; Parlikad, A.K.; et al. Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. *Internet Things Cyber-Phys. Syst.* **2024**, *4*, 19–23. [[CrossRef](#)]
38. Yu, H. The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. *Heliyon* **2024**, *10*, e24289. [[CrossRef](#)] [[PubMed](#)]
39. Barton, K.C.; Levstik, L.S. *Teaching History for the Common Good*; Routledge: London, UK, 2004.
40. Rothery, J.; Stenglin, M. Entertaining and instructing: Exploring experience through story. In *Genre and Institutions: Social Processes in the Workplace and School*; Christie, F., Martin, J.R., Eds.; Continuum: London, UK, 1997; pp. 231–263.
41. Gardner, J. *The Art of Fiction: Notes on Craft for Young Writers*; Vintage: New York, NY, USA, 2010.
42. Swedish Research Council. *Good Research Practice*; Swedish Research Council: Stockholm, Sweden, 2017.
43. Alkaiissi, H.; McFarlane, S.I. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus* **2023**, *15*, 1–4. [[CrossRef](#)] [[PubMed](#)]
44. Endacott, J.L.; Brooks, S. Historical empathy: Perspectives and responding to the past. In *The Wiley International Handbook on History Teaching and Learning*; Metzger, S.A., McArthur Harris, L., Eds.; Wiley Blackwell: Hoboken, NJ, USA, 2018; pp. 203–225.
45. Bertram, C.; Weiss, Z.; Zachrich, L.; Ziai, R. Artificial intelligence in history education. Linguistic content and complexity analyses of student writings in the CAHisT project (Computational assessment of historical thinking). *Comput. Educ. Artif. Intell.* **2021**, *10038*, 1–48. [[CrossRef](#)]
46. Brookbanks, G. Teaching and learning of History at a high school level—The reality of AI/ChatGPT and the process of assessing understanding. *Yesterday Today* **2023**, *29*, 120–123. [[CrossRef](#)]
47. Levstik, L.S. The relationship between historical response and narrative in a sixth-grade classroom. *Theory Res. Soc. Educ.* **1986**, *14*, 1–19. [[CrossRef](#)]

48. VanSledright, B.; Brophy, J. Storytelling, imagination, and fanciful elaboration in children's historical reconstructions. *Am. Educ. Res. J.* **1992**, *29*, 837–859. [[CrossRef](#)]
49. Karlsson, P.A. *Undervisning och Lärande i Historia: Ett Kreativt rum för Narrativ Kompetens*; Acta Universitatis Stockholmiensis: Stockholm, Sweden, 2014.
50. Christie, F.; Derewianka, B. *School Discourse: Learning to Write across the Years of Schooling*; A&C Black: London, UK, 2010.
51. D'Adamo, L.; Fallace, T. The multigenre research project: An approach to developing historical empathy. *Soc. Stud. Res. Pract.* **2011**, *6*, 75–88. [[CrossRef](#)]
52. Endacott, J.L. Reconsidering affective engagement in historical empathy. *Theory Res. Soc. Educ.* **2010**, *38*, 6–47. [[CrossRef](#)]
53. Berg, M.; Persson, A. The Didactic Function of Narratives: Teacher discussions on the use of challenging, engaging, unifying, and complementing narratives in the history classroom. *Hist. Encount.* **2023**, *10*, 44–59. [[CrossRef](#)]
54. Michel-Villarreal, R.; Vilalta-Perdomo, E.; Salinas-Navarro, D.E.; Thierry-Aguilera, R.; Gerardou, F.S. Challenges and opportunities of generative AI for higher education as explained by ChatGPT. *Educ. Sci.* **2023**, *13*, 856. [[CrossRef](#)]
55. Sibilin, C.S. Education and the Epistemological Crisis in the Age of ChatGPT. *Crit. Rev.* **2023**, *35*, 414–425. [[CrossRef](#)]
56. McIntire, A.; Calvert, I.; Ashcraft, J. Pressure to Plagiarize and the Choice to Cheat: Toward a Pragmatic Reframing of the Ethics of Academic Integrity. *Educ. Sci.* **2024**, *14*, 244. [[CrossRef](#)]
57. Biesta, G. Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educ. Assess. Eval. Account.* **2009**, *21*, 33–46. [[CrossRef](#)]
58. Heimans, S.; Biesta, G.; Takayama, K.; Kettle, M. ChatGPT, subjectification, and the purposes and politics of teacher education and its scholarship. *Asia-Pac. J. Teach. Educ.* **2023**, *51*, 105–112. [[CrossRef](#)]
59. AI and Humanism: A Conversation with ChatGPT. Available online: <https://hapihumanist.org/2023/05/20/a-conversation-with-chatgpt/> (accessed on 2 April 2024).

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.