ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Artificial intelligence in Higher Education: Evaluating a Custom GPT for Teaching Planning in a Course on American History

Inteligencia artificial en la educación superior: evaluación de un GPT personalizado para la planificación docente en una asignatura de Historia de América

Antonio Carrasco-Rodríguez
University of Alicante
antonio.carrasco@ua.es
0000-0003-3913-0200

Recibido: 19/03/25
Aceptado: 20/12/25

Resumen
Este estudio evalúa las aplicaciones de un asistente de ChatGPT para la planificación docente y la preparación de contenidos de una asignatura universitaria sobre la Historia de América. El asistente fue diseñado para ayudar en tareas, como la creación de contenidos teóricos y actividades prácticas, el entrenamiento del alumnado y la evaluación. Metodológicamente, combina un enfoque iterativo con la evaluación cuantitativa y cualitativa de 10 usos del asistente. Los resultados destacan el potencial del asistente para ahorrar tiempo, mejorar los materiales docentes y optimizar procesos de enseñanza y aprendizaje. No obstante, también presenta limitaciones, como la falta de profundidad en algunos contenidos, la recomendación de bibliografía incorrecta y la necesidad de una supervisión constante por el docente. El estudio concluye que la inteligencia artificial generativa tiene un gran potencial como herramienta de apoyo para los profesores y subraya la necesidad de que se realicen más investigaciones empíricas en este campo.

Palabras clave
Inteligencia artificial, enseñanza superior, historia americana, aprendizaje, evaluación

Abstract
This study evaluates the applications of a ChatGPT assistant for teaching planning and content preparation in a university course on American History. The assistant was designed to support tasks such as the creation of theoretical content and practical activities, student training, and assessment. Methodologically, the study combines an iterative approach with a quantitative and qualitative evaluation of 10 specific uses of the assistant. The results highlight its potential to save time, enhance teaching materials, and optimize teaching and learning processes. However, limitations were also identified, including a lack of depth in some content, inaccurate bibliography recommendations, and the need for constant teacher supervision. The study concludes that generative artificial intelligence has significant potential as a support tool for educators and underscores the need for further empirical research in this field.

Keywords
Artificial intelligence, higher education, American history, learning, evaluation

To cite this paper: Carrasco-Rodríguez, Antonio (2026). Inteligencia artificial en la educación superior: evaluación de un GPT personalizado para la planificación docente en una asignatura de Historia de América. *Panta Rei. Revista Digital de Historia y Didáctica de la Historia, 20*. DOI: 10.6018/pantarei.653731

1. Introduction

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Since 2023, generative artificial intelligence (AI) has had a significant impact on a wide range of sectors, including education (García-Peñalvo, 2024). Large Language Models (LLMs) have demonstrated their capacity to generate texts, perform diverse tasks, and adapt to specific needs, opening new possibilities for both educators and students (Garrido-Merchán et al., 2023). As a result, academic interest in the educational applications of AI has grown rapidly, alongside increasing concern about its potential risks and unintended consequences.

Despite this growing body of literature, research on the use of generative AI in educational contexts remains uneven. Many studies have focused on theoretical reflections, ethical debates, institutional policies, or the perceptions and attitudes of students and teachers, while empirical analyses of concrete instructional uses are still relatively limited (Labadze et al., 2023; Idroes et al., 2023; McGrath et al., 2023; Sánchez, 2023; Sok & Heng, 2023). This imbalance is particularly evident in higher education, where the integration of generative AI into tasks such as teaching planning, the creation of educational materials, and student assessment requires more contextualized and practice-oriented investigation (Almasri, 2024; Dogan et al., 2023).

Within the Humanities, and especially in the field of History, this gap becomes even more pronounced. Teaching History entails specific epistemological and didactic challenges, including the development of historical thinking, the critical interpretation of narratives, and the careful handling of historiographical debates and sources. These characteristics raise legitimate concerns regarding the use of generative AI, particularly with respect to accuracy, bias, simplification, and the reproduction of dominant or eurocentric narratives. At the same time, they also make History a particularly suitable field for exploring the pedagogically mediated use of generative AI tools under conditions of systematic human supervision.

Against this background, there is a clear need for empirical, use-oriented studies that document how generative AI tools are employed in History teaching, how their outputs are evaluated, and what strengths and limitations emerge from their sustained use in real instructional settings. This article seeks to address this gap by presenting an evaluative case study on the use of a custom GPT assistant based on ChatGPT for teaching planning and the preparation of educational content in a university course on Pre-Columbian and Colonial Hispanic American History. Rather than focusing on perceptions or hypothetical scenarios, the study analyzes ten specific instructional uses of the assistant related to teaching planning, materials creation, student training, and assessment support, assessing their pedagogical effectiveness, efficiency, and limitations within a framework of continuous instructor oversight.

2. Theoretical Framework

The scientific literature on generative artificial intelligence in education has expanded rapidly in recent years (Ayeni et al., 2024; Gökçearslan et al., 2024; Onesi-Ozigagun et al., 2024; Sperling et al., 2024). However, a substantial proportion of this research consists of theoretical analyses or general reviews addressing the capabilities, opportunities, and risks associated with generative AI technologies (Labadze et al., 2023; Sok & Heng, 2023). Empirical studies examining the concrete integration of generative AI

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

into teaching and learning processes remain comparatively scarce, particularly in higher education contexts (Almasri, 2024; Dogan et al., 2023).

Research focusing specifically on the educational use of generative AI in History is even more limited. As of early 2025, only a small number of scientific publications have addressed this topic, with the majority consisting of conceptual reflections on the potential benefits and risks of AI in History education. A smaller group of studies has explored students' and teachers' perceptions of AI tools (Carrasco-Rodríguez, 2024b; Carrasco-Rodríguez et al., 2024; Fernández-Arrillaga et al., 2024; Lazareva et al., 2024; Rosser-Limiñana & Soler-Ortiz, 2024), while an even more restricted set of works has adopted an empirical approach, presenting pedagogical proposals, case studies, or evaluations of specific teaching applications (Acun & Acun, 2023; Bertram et al., 2021; Calderón Pantoja, 2024; Carrasco-Rodríguez, 2023, 2024a; Carretero & Gartner, 2024; Fareed et al., 2024; Hutson, 2024; Kindenberg, 2024; Melero-Muñoz, 2023; Mottl & Musílek, 2024; Soler-Ortiz & Rosser-Limiñana, 2024a, 2024b; Tirado-Olivares et al., 2023).

From a didactic perspective, History is characterized by specific epistemological demands that distinguish it from other disciplines. Teaching and learning History involve more than the transmission of factual information; they require the development of historical thinking, including the analysis of causality, change and continuity, the evaluation of sources, and the construction and critique of historical narratives. These features make the uncritical use of generative AI particularly problematic, as AI-generated outputs may privilege simplified explanations, dominant historiographical perspectives, or implicit biases embedded in training data.

At the same time, these same characteristics create opportunities for the pedagogically mediated use of generative AI as a support tool. When integrated into teaching practice under conditions of explicit instructional design and systematic human oversight, generative AI can assist educators in planning courses, organizing content, generating learning materials, and designing assessment instruments, while also serving as an object of critical analysis for students. In this regard, ChatGPT is of particular interest due to its custom GPT assistants, which allow creators to define specific instructions and incorporate carefully selected knowledge sources to guide model behavior and outputs.

To date, however, the use of custom GPT assistants has not been the subject of systematic analysis within the field of History education. Their capabilities, limitations, and pedagogical value as support tools for educators remain largely unexplored. The present study addresses this gap by situating the evaluation of a custom GPT assistant within the broader debates on generative AI in education, historical thinking, and responsible pedagogical mediation, providing an empirical contribution grounded in sustained instructional practice.

3. Methodology

3.1. Objectives

This article aims to explore and evaluate the usefulness of ChatGPT, through a custom GPT assistant, for university History instructors in teaching planning and content preparation. The study is based on the hypothesis that its use could revolutionize the daily work of

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

educators in this discipline by saving time on routine tasks, helping them expand their subject knowledge, improving the didactic quality of their content, and optimizing their evaluation processes.

To achieve this objective and address the hypothesis, the study focused on planning and preparing the course America: History from Colonization to the Present, taught in the fourth year of the undergraduate degree in History at the University of Alicante, Spain. Specifically, the study concentrated on thematic blocks dedicated to the History of Pre-Columbian America and Colonial Hispanic America. The novelty of this work lies in its empirical approach, which may serve as a foundation for conducting similar studies in other courses, educational levels, and disciplines.

3.2. Study Context

This study aims to analyze the impact of a custom GPT assistant on the planning of the course America: History from Colonization to the Present. The assistant was designed to perform tasks such as reviewing the course syllabus, creating lecture notes and supplementary materials with critical analysis, supporting autonomous student training, designing practical activities and exams, and proposing evaluation criteria and rubrics. Its implementation followed an iterative approach, where collaboration between the instructor and the assistant enabled optimization of the results obtained for each task.

The course America: History from Colonization to the Present is part of the undergraduate degree in History curriculum at the University of Alicante and is divided into two sections. The first section, which is the primary focus of this analysis, covers Pre-Columbian and Colonial Hispanic American history and is taught by the author of this study. The second section, taught by a Modern History professor, focuses on the history of the Americas from the independence movements to the present.

For the 2024–2025 academic year, the course had an enrollment of 97 students in their final year of study. These students had prior training in European and Spanish Early Modern History but, with few exceptions, lacked intermediate or advanced knowledge of American History. Additionally, they possessed intermediate to upper-intermediate proficiency in the use of generative artificial intelligence and custom GPT assistants, a skill set developed in courses taught by the same instructor over the past two academic years. The course was delivered in a face-to-face format, employing a combination of lectures and practical activities. Classroom sessions were complemented by independent tasks designed to strengthen students' critical and analytical abilities, encouraging active participation in the learning process.

3.3. Uses of the GPT Assistant

The custom GPT assistant for this course was designed to be utilized by both students and instructors. Specifically, for instructors, the assistant is equipped to assist with tasks related to planning, teaching delivery, evaluation, and even administrative management. This study focuses on the potential applications of the assistant and generative AI in teaching planning, broadly defined. To this end, ten specific uses of the assistant were defined and systematically explored to evaluate its efficiency and utility: reviewing the course syllabus, designing practical activities, creating lecture notes, evaluating theoretical content,

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

reviewing and enhancing the bibliography, creating supplementary materials, supporting autonomous student training, designing exams, selecting evaluation criteria for assignments and creating rubrics, and selecting evaluation criteria for theoretical exams and creating rubrics.

3.4. Evaluation Procedures for the Uses of the GPT Assistant

The evaluation of the custom GPT assistant was conceived as a use-oriented and context-sensitive process, consistent with the exploratory and applied nature of the study. Rather than relying on a single uniform metric, the evaluation strategy combined quantitative ratings and qualitative analysis, adapted to the specific objectives and characteristics of each instructional use examined.

Quantitative evaluation through a five-point Likert-type scale (1–5) was applied selectively, only in those cases where comparative assessment across multiple outputs was methodologically appropriate. Specifically, numerical ratings were employed in Use 6 (Proposal and Creation of Supplementary Materials) and Use 7 (Autonomous Student Training), where the assistant generated alternative materials or activities that could be meaningfully compared. In these uses, the scale assessed overall performance according to explicit criteria defined in the corresponding sections, including curricular adequacy, historical rigor, didactic quality, clarity, and practical usefulness.

Within this framework, a score of 1 indicated outputs that were inadequate or unusable without substantial revision; a score of 3 referred to outputs that were partially valid but required significant improvement or contextualization; and a score of 5 denoted outputs that met the established criteria to a high standard and could be incorporated into teaching practice with minimal modification. Intermediate values were assigned when outputs fell between these reference points. Numerical ratings were not interpreted in isolation but served as a synthetic indicator supporting broader qualitative judgment.

Qualitative analysis constituted the core of the evaluation across all ten instructional uses. For each use, the assistant's outputs were examined through detailed qualitative review, focusing on the accuracy and depth of historical content, coherence with course objectives, alignment with current historiographical approaches, didactic effectiveness, and inclusivity. These analyses were documented through structured analytical notes, which informed both the iterative refinement of prompts and the final evaluative judgments reported in the results.

Evaluation followed an iterative workflow involving repeated cycles of prompt formulation, output generation, and instructor review. In tasks related to content creation and revision—particularly the development of lecture notes and supplementary materials—AI-generated outputs were systematically contrasted with established teaching manuals and with recent historiographical scholarship published within the previous eight years. Iterations continued until the materials met predefined standards of historical rigor, pedagogical quality, and inclusivity, ensuring that evaluation addressed both the quality of final outputs and the effectiveness of the assistant as a supervised support tool.

Finally, although the primary focus of the evaluation was on instructor-mediated uses of the assistant, informal student feedback was considered as complementary qualitative

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

evidence. Students interacted with the assistant during the course to support practical assignments and exam preparation, and their feedback, collected during classroom working sessions, was generally positive—particularly regarding the assistant's capacity to generate summaries, outlines, and comparative tables. This feedback was not treated as a formal perception study but as exploratory triangulation supporting the interpretation of the evaluation results.

3.5. Description of the custom GPT assistant

This study employed a single custom GPT assistant created within ChatGPT using the Custom GPTs (GPT Builder) environment, specifically designed for the undergraduate course America: History from Colonization to the Present. The assistant was conceived as a course-specific support tool to assist both the instructor's teaching planning and content preparation tasks and students' learning processes (study support and completion of practical assignments). It was created in October 2024 and iteratively refined by the author between October and December 2024, before being used throughout the course delivery period (January–June 2025) and during the resit examination phase (July 2025).

The assistant's core configuration was defined through a set of system-level instructions that established its pedagogical role and operational constraints. These instructions specified: (a) its function as a discipline- and course-specific assistant for History of America; (b) the differentiation between two user profiles (instructor and students), with distinct communicative goals, depth of explanation, and pedagogical support; (c) the prioritization of course-approved materials over general model knowledge when responding to queries; (d) multilingual interaction, responding in the language used by the user; and (e) the use of inclusive and respectful language. The scope of tasks covered by the assistant was explicitly aligned with the ten teaching and assessment uses analyzed in this study.

To ground the assistant's outputs in validated academic content, a dedicated knowledge base was incorporated. This knowledge base consisted exclusively of course-specific materials in Spanish, including updated lecture notes, slide presentations, the official course syllabus, and supporting instructional documents. This design aimed to reduce reliance on generic model knowledge and to ensure that generated outputs were consistent with the course's learning objectives, historiographical orientation, and assessment criteria.

Regarding technical parameters, the GPT Builder environment does not provide direct access to low-level generation settings such as temperature, top-p sampling, or explicit token limits, nor does it allow the export of run-level logs. Consequently, this study reports only those configuration elements that are directly controllable within the platform (system instructions, knowledge base, and enabled capabilities). Control over output specificity, rigor, and pedagogical alignment was therefore implemented indirectly through carefully designed prompts and systematic iterative refinement under human supervision. While this limits strict computational reproducibility, it reflects the practical constraints of deploying commercial generative AI tools in authentic educational settings and is consistent with the exploratory and evaluative aims of this study.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Finally, the assistant was used as a support tool rather than as an autonomous agent. All outputs intended for teaching or assessment purposes—such as lecture materials, activity guides, rubrics, and evaluation instruments—were systematically reviewed, validated, and edited by the instructor prior to use. The workflow followed iterative cycles of prompting, output evaluation, and revision, which functioned both as a quality assurance mechanism and as a means of progressively aligning the assistant's responses with the pedagogical, epistemological, and ethical goals of the course. This human-in-the-loop approach was a central design principle underpinning the responsible use of generative AI in this educational context.

4. Results

To facilitate a transversal reading of the ten instructional uses analyzed in this study, Table 1 synthesizes their pedagogical purposes, task types, evaluation criteria, metrics, main outcomes, and key limitations.

Table 1
Synthesis of the ten instructional uses of the custom GPT assistant

N°	Instructional use	Purpose & task	Evaluation approach	Main outcome	Key limitation
1	Syllabus review	Curriculum coherence; evaluation and revision of the course syllabus	Qualitative review (coherence and alignment with objectives and assessment criteria)	Improved internal consistency and clarity of the syllabus	Weak performance in providing curated and reliable external links
2	Practical activity design	Active learning; generation of practical activities (gamification, role-play, and GPT-supported tasks)	Qualitative review (pedagogical relevance, feasibility, and alignment with course design)	High creativity and generation of usable activity proposals with supporting documentation	Requires contextual adaptation to cohort characteristics and time constraints
3	Lecture notes development (incl. structuring)	Content creation and coherence; structuring and drafting lecture notes	Qualitative review (rigor, coverage, clarity, didactic quality) and time/volume indicators	Major efficiency gains and substantial improvement of teaching materials (e.g., 253 pages produced in 10 days)	Occasional lack of analytical depth and examples; imperfect use of uploaded sources; minor technical issues
4	Content evaluation (lecture notes/texts)	Quality control; critical evaluation of instructional texts	Qualitative assessment (accuracy, coherence, depth, bias detection, and improvement value)	Effective identification of errors, inconsistencies, and potential biases, with useful feedback	Reviews not always exhaustive for highly complex historical issues

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

N°	Instructional use	Purpose & task	Evaluation approach	Main outcome	Key limitation
5	Bibliographic support	Research support; recommendation of key academic works	Quantitative indicators (% of verified vs. fabricated references) combined with qualitative relevance checks	Higher validity and usefulness of English-language recommendations	Very high fabrication or non-verifiability rate in Spanish-language references; linguistic and canonical bias
6	Supplementary materials	Didactic enrichment; design and generation of complementary textual and visual materials	Likert-type scale (1–5) supported by qualitative analytical notes	Strong multimodal proposals, high creativity, and rapid production of complementary resources	Requires instructor adaptation and systematic source verification
7	Autonomous student training	Comparative historical thinking; support for self-study and review	Likert-type scale (1–5) supported by qualitative analytical notes	Effective study support through summaries, outlines, and comparative tables	Risk of oversimplification; requires instructor guidance to ensure nuance
8	Exam design	Assessment planning; generation and redesign of examination materials	Quantitative indicators (% of valid questions) combined with qualitative review	High efficiency and usefulness in generating question banks and exam structures	Requires human validation and calibration to ensure appropriate difficulty
9	Rubric design (advanced prompt task)	Competence assessment; design of analytic rubrics for complex student tasks	Qualitative review (clarity, applicability, alignment with learning outcomes)	Robust and practical evaluation frameworks that clarify assessment criteria	Weighting of criteria remains instructor-dependent
10	Exam evaluation support	Assessment design; definition of criteria and rubrics for different exam question types (short-answer, outline, and essay)	Qualitative review of proposed criteria and weightings	Rapid and effective generation of coherent evaluation criteria and rubrics for multiple exam formats, requiring only minor adjustments	Final weighting decisions and application of criteria remain dependent on instructor judgment

Note: Table created by the author based on the content of the 10 uses of the custom GPT assistant. Note: Likert-type ratings (1–5) were applied selectively (Uses 6 and 7). Other instructional uses were evaluated primarily through qualitative review or use-specific quantitative indicators. Representative prompts and detailed evaluation criteria are provided in the Supplementary Material.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

4.1. Use 1: Reviewing the Course Syllabus

The assistant performed four tasks, which included reviewing the course description, objectives, and evaluation criteria, as well as providing relevant links.

Regarding the course description, the assistant first revised the existing text, significantly enhancing it to the extent that, with minor adjustments, it could be included in the official syllabus. Subsequently, it generated a completely new description without referencing the original text, producing an even better version. This new text offered a broader conceptual scope, highlighted the course's relevance, emphasized competencies such as critical and interdisciplinary analysis, and its application in educational settings, while maintaining an inclusive and academic tone.

A similar approach was followed to review the course objectives, both general and specific. The assistant was asked to refine the existing objectives and to create new ones from scratch. In both cases, its drafts outperformed the current text, with the newly created objectives demonstrating particularly high quality. These new objectives featured a clearer and more logical structure, deeper analytical depth, a more explicit connection to practical skills in research and teaching, and a critical perspective aligned with the contemporary historical context.

For the evaluation criteria, the assistant proposed improvements to the text's organisation by restructuring it into lists and enhancing the language to make it clearer and more motivational. This approach eliminated the punitive tone of the original version, instead presenting the criteria in a more encouraging and constructive manner.

Finally, the assistant was tasked with compiling a list of high-quality, reliable, and verified links related to Pre-Columbian history, the colonial period, the independence movements, and Modern Latin America. The instructions explicitly emphasized avoiding Wikipedia and ensuring academic rigor. Unfortunately, the results were disappointing, as the links provided were neither accurate nor reliable.

4.2. Use 2: Creation of Practical Activities

The assistant proposed practical activities aligned with the course's learning objectives, designed to promote creativity, critical analysis, and the development of advanced digital skills. Once the activities were selected, the assistant helped to develop and plan their implementation. Two types of activities were requested: those intended to assess classroom participation and those designed as practical assignments to be completed autonomously by students in pairs.

To evaluate participation, the assistant was instructed to create 12 tasks to be carried out with the support of the course assistant during 30-minute sessions over 12 different days. The content to be covered in each session was provided, along with five suggested activity types: comparative analysis, study through different historiographical perspectives, connection with contemporary issues, creation of first-person texts, and role-play involving problem-solving within assigned roles.

The assistant proposed 12 types of activities, of which seven were accepted: creation of a gamified activity, case study analysis, debate simulation, decision-making simulation,

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

critical argumentation, artistic creation using artificial intelligence, and historical scenario construction. The assistant then created activities for each session, ensuring that tasks were not repeated and were tailored to the themes of the classes. For example, in session 4, which focused on the Postclassic Andean period, the assistant proposed the creation of a first-person narrative describing the daily life of an aclla. For session 8, on the conquest of the New World, the assistant suggested a case study analyzing the fall of the Mexica Empire from military, political, economic, and social perspectives.

On the other hand, to evaluate the practical component of the course, the assistant was asked for ideas, specifying that the assignments had to be completed in pairs, require approximately 10 hours of independent work per student, and involve the use of the course assistant. The GPT provided five initial proposals, which were not directly useful. However, they became the foundation for an idea that eventually developed into the practical assignment for the course. The students were tasked with creating a ChatGPT prompt that would allow them to converse with a character from the colonial period, executed within the course assistant for better contextual understanding.

The assistant was then asked to further develop the idea and create a guide for completing the assignment. In the guide, it included the title of the assignment, the objectives of the activity, and instructions organized in steps, covering aspects such as selecting the character, conducting research, designing the prompt, testing and adjusting it, and writing the final report. It also specified the evaluation criteria and their respective weightings: historical accuracy (40%), creativity (20%), the quality of the prompt (20%), and the quality of the report (20%).

The assistant's contribution to creating the practical activities was excellent. It demonstrated the ability to generate interesting, useful, pedagogical, and relevant proposals, to adapt to the instructor's needs and requirements, and to assist in the design, planning, and implementation of the activities. Additionally, it saved time and enriched the instructor's ideas.

4.3. Use 3: Creation of New Lecture Notes

This use posed the greatest challenge for the course assistant. The Pre-Columbian and Colonial Hispanic American History modules consist of eight topics:

1. Before It Was America.
2. European Expansion and the New World.
3. Law and Justice in the Genesis of the New Society.
4. Population and Economy.
5. Administrative Structures and Society.
6. The Bourbon Reforms.
7. Women in Colonial Hispanic America.
8. The Church in the New World.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

The lecture notes were originally created eight years ago. Although small modifications had been made to the topics over time, the course instructors had been considering rewriting them for several academic years to expand their content and incorporate the latest scientific advances in historiography. These tasks were finally undertaken with the support of the assistant, following a step-by-step process.

To begin, the assistant was asked to specify the table of contents for each topic. It was tasked with both creating a new structure from scratch and improving the organization of the existing notes, which had been included in its specific knowledge base. Once the table of contents was finalized, the content creation phase began. Each section or subsection was developed using one of two strategies: in some cases, the assistant generated an initial draft from scratch, which was then supplemented by the instructor's contributions; in others, it enhanced the instructor's original notes. The first approach proved more effective for general texts, while the second was better suited for specialized content that required greater depth and nuance.

After drafting the text, a dual verification process was implemented to ensure accuracy. The assistant conducted an initial review to verify the accuracy of the content and justify its claims, which proved highly beneficial. This was followed by a thorough review by the instructor to validate the information and correct any inconsistencies. Following the verification process, the assistant was asked to propose content improvements. It integrated these enhancements while marking the changes clearly to facilitate review and validation. This iterative approach helped refine the quality of the materials further.

Once the text was polished, a final review was conducted to enhance the writing, organization, and overall style. The assistant played a role in refining clarity and coherence, ensuring that the lecture notes maintained an academic and pedagogically sound tone. Lastly, the assistant was tasked with proposing specific bibliographic references for each section or subsection of the notes. It also provided a brief commentary on the relevance of each reference, although its effectiveness in recommending accurate and high-quality sources varied.

The overall outcome of this process was highly positive. In approximately 80 hours, spread over 10 days, the new versions of the lecture notes for all 8 topics were created with the support of the GPT assistant. The updated notes consisted of a total of 253 pages (110 more than the original notes), formatted in 11-point Aptos font with 1.0 line spacing.

Regarding the evaluation of the process, the assistant performed well in creating tables of contents, both from scratch and by improving existing structures. However, it encountered some difficulties when generating outlines with more than three levels of depth.

In content creation, the assistant excelled in several areas. It demonstrated high-quality writing, employed language and style appropriate for an academic and pedagogical tone, and effectively established connections with other topics. It also successfully incorporated diverse and minority perspectives and explained relationships between social groups and their medium- and long-term impacts. Additionally, the assistant stood out for including geographical, interdisciplinary, and popular aspects, as well as for relating the texts to global contexts.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Nevertheless, the assistant faced challenges with depth and integration of examples in its writing, occasional inaccuracies, and minimal yet notable Eurocentric biases. It was more effective when dealing with general topics and less so with more specific issues.

In terms of accuracy verification, the assistant performed well. It conducted thorough analyses of the texts, provided detailed explanations about the accuracy or lack of rigor in the content, and offered recommendations for error correction. However, its effectiveness in proposing improvements was limited for two reasons. First, it predominantly relied on English-language bibliographic sources, often overlooking works written in Spanish. Second, it encountered technical issues when creating comments within the canvas and struggled to directly implement changes.

The assistant performed adequately in making recommendations to improve text organization, writing, and style. However, it again struggled to directly apply these improvements to the canvas. Regarding bibliographic suggestions, as noted earlier, it mostly provided sources in English.

Finally, while the canvas system was a valuable tool, it had some limitations. These included difficulties in directly integrating changes into the text, challenges with automatically detecting manual edits, instances of content being deleted or summarized without notice upon exceeding the canvas size limit, and slowdowns in content creation as the chats for each topic grew.

The final assessment of this use is very positive. Despite its limitations, the assistant enabled the completion of a task of exceptional quality within 10 days of nearly exclusive dedication, a task that likely would have required over a month without the support of the GPT assistant and generative artificial intelligence.

4.4. Use 4: Evaluation of Course Content

Once the lecture notes were completed, they were evaluated with the support of the GPT assistant. A systematic analysis procedure was developed, consisting of 16 successive steps designed to identify the strengths and weaknesses of the new notes and to make specific improvement proposals, facilitating their integration into the texts. These are the 16 evaluation criteria:

1. Initial rapid evaluation (identification of issues, verification of alignment with objectives and competencies, and prioritization of revisions).

2. Structure review (logical sequencing of sections, appropriateness of titles, and identification of redundancies between sections).

3. Analysis of historical content (accuracy, depth, and identification of gaps or omissions).

4. Cross-sectional connections (adequate relationships with other topics and an integrated and consistent approach).

5. Analysis of first- and second-order historical concepts.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

6. Bias detection (identification of Eurocentric perspectives, commentary on controversial aspects, and evaluation of balanced representation).

7. Inclusion of diverse perspectives (indigenous, Afro-descendant, and gender perspectives, and marginalized historical actors).

8. Depth in cultural interactions (explanation of cultural exchanges among Europeans, Amerindians, and Afro-descendants; mutual transformations; and examples of cultural and religious syncretism).

9. Evaluation of long-term impacts (connections with the present, influence of historical processes on identity formation, and reflection on cultural, political, social, and economic legacies).

10. Inclusion of comparisons and global contexts (relations between processes in Colonial Hispanic America and global events/dynamics, and comparisons with other colonizations in Asia, Africa, and the Americas).

11. Language and style evaluation (clarity, objectivity, accessibility, inclusivity, and avoidance of subjective or biased judgments).

12. Historiographical updates (recent research, current historiographical debates, and replacement of outdated references with more reliable or recent ones).

13. Review of geographic inclusiveness (representation of all areas of the Americas, including less-studied regions).

14. Interdisciplinarity (incorporation of perspectives from anthropology, sociology, economics, or auxiliary sciences of history).

15. Critical evaluation of sources (assessment of source quality and identification of gaps).

16. Inclusion of oral and popular narratives (references to indigenous and Afro-descendant testimonies, and inclusion of popular traditions and collective memories).

To enhance the quality of the topics, all the newly created notes were evaluated following this procedure. The assistant demonstrated both strengths and weaknesses.

Among the positive aspects, the assistant provided a substantial amount of information on the content. It generated multiple improvement proposals for each evaluation criterion, most of which were interesting, accurate, and well-founded. Additionally, the assistant facilitated their implementation by providing precise information about the location of the content to be modified and the specific texts to be replaced. Its suggestions related to inclusivity, interdisciplinarity, connections with the present, and representativity were particularly outstanding, adding significant value to the topics.

However, the assistant also exhibited certain limitations. For some criteria, its analysis, while detailed, was not exhaustive, and its proposals occasionally lacked examples. In some areas, certain criteria were not applicable. Lastly, the assistant's iterative capability

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

to generate new improvement proposals created the potential for the review process to become endless.

4.5. Use 5: Bibliography

The GPT assistant's applications related to bibliography were somewhat better than expected but still insufficient compared to other uses of artificial intelligence. The assistant was tasked with providing bibliographic references for three purposes: to add solidity and accuracy to the lecture notes, to offer specific references for each section of the topics, and to create a general bibliography for each topic. In all cases, the assistant was asked to include a brief explanation of each reference and a justification for its selection.

Based on prior negative experiences with bibliographic requests, the assistant was instructed that all references must be real, not fabricated, relevant, verified, locatable in major specialized databases (such as Web of Science, Scopus, Google Scholar, or Dialnet), and authoritative sources. This aimed to minimize the errors generative artificial intelligence often introduces in such tasks.

The results had more weaknesses than strengths. In general, the assistant tended to present bibliographic references in English, likely due to the prevalence of English-language materials in its training data. It provided very few references in Spanish, which limited the quality and quantity of its recommendations, as Spanish-language sources dominate historiography on Pre-Columbian and Colonial Hispanic America. Of the recommendations in English, 85% were real, verified, and authoritative. However, nearly 90% of the recommendations in Spanish were fabricated. This reaffirmed the already acknowledged necessity of verifying every bibliographic recommendation. While the assistant made interesting improvement suggestions based on English sources, manually searching for references in specialized catalogs and databases remains much more efficient.

On the other hand, the assistant proved useful for formatting references in APA (7th edition). It was able to convert bibliographic entries from various standards into APA format with minimal errors.

4.6. Use 6: Proposal and Creation of Supplementary Materials

With the lecture notes and bibliographies completed, the next step was to enrich the course content with supplementary materials. The assistant was provided with an initial list of nine categories of materials. It generated 20 additional proposals, 11 of which were incorporated into the final list. The assistant was then asked to create five materials for each category.

To evaluate the quality of these materials, the assistant was tasked with proposing three evaluation criteria for each category and five generic criteria for an overall assessment. Below is the list of supplementary material types (including the original nine and the eleven recommended by the assistant), the evaluation criteria, and the average ratings of the content created by the GPT (rated on a scale from 1 to 5, where "1" represents "Very Unsatisfactory" and "5" represents "Very Satisfactory"):

1. Anecdotes and curiosities. Criteria: relevance, engagement, and accuracy. Score: 4.3.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

2. Content expansions for lecture notes. Criteria: depth, connection, and clarity. Score: 4.

3. Connections to the present. Criteria: pertinence, timeliness, and reflective impact. Score: 5.

4. People, events, or processes that contributed to a better world or had a positive impact on history. Criteria: inspiration, balance, and relevance. Score: 4.7.

5. Learning history through games, video games, movies, series, documentaries, artworks, literature, and comics. Criteria: analysis, pedagogical value, and bias detection. Score: 5.

6. Controversial topics. Criteria: neutrality, promotion of debate, and rigor. Score: 5.

7. Historiographical debates. Criteria: representation of perspectives, argumentative clarity, and rigor. Score: 3.3.

8. Diverse perspectives. Criteria: inclusivity, originality, and rigor. Score: 5.

9. Linking history with other disciplines. Criteria: connection with other fields, complementarity, and clarity. Score: 5.

10. Alternative narratives. Criteria: unexplored perspectives, coherence, and educational value. Score: 5.

11. Life stories or biographies. Criteria: humanization, variety, and contextualization. Score: 5.

12. Analysis of objects and artifacts. Criteria: description, contextualization, and interpretation. Score: 5.

13. Case studies. Criteria: focus, evidence, and connection. Score: 5.

14. Histories of words or expressions. Criteria: etymology, relevance, and narrative appeal. Score: 4.7.

15. Annotated quotes. Criteria: context, relevance, and selection of the quote. Score: 3.3.

16. Ethical questions about historical dilemmas. Criteria: reflective capacity, absence of bias, and clarity. Score: 5.

17. Cultural glossaries. Criteria: accuracy, relevance, and variety. Score: 5.

18. Impacts of science and technology. Criteria: pertinence, timeliness, and interdisciplinarity. Score: 5.

19. City profiles. Criteria: context, description, and cultural impact. Score: 5.

20. Historical travel guides. Criteria: usability, variety, and pedagogical value. Score: 4.7.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

In summary, 13 of the 20 types of supplementary materials received the highest rating. The overall average score for all categories was very high (4.70), with the lowest score being 3.30 for historiographical debates and annotated quotes, and 4.0 for content expansions. The global evaluation criteria recommended by the assistant were thematic relevance (5), academic rigor (4), clarity (5), interest (5), and educational impact (5).

The assistant performed brilliantly in these tasks. It made valid proposals for the categories, generating a wide variety and quantity of content very quickly, with high levels of accuracy, appeal, and interest. The materials were especially valuable in categories such as connections to the present, human rights advocacy, respect for diversity, balanced representation, marginalized perspectives, alternative learning materials, and interdisciplinary contributions.

Furthermore, the approach to materials dealing with controversial topics or ethical questions was appropriate—neutral when required and free from bias. However, some limitations were noted, including a lack of depth in some cases, a preference for generic topics (although the assistant was almost always able to create more specific content when requested), difficulties in selecting suitable sources for historiographical debates, and the fabrication of quotes from primary sources.

4.7. Use 7: Autonomous Student Training

Another interesting application of the GPT assistant is its use by students for independent learning. Since it incorporates the course lecture notes in its specific knowledge base, the custom GPT can create review materials and activities that students can complete and have corrected by the assistant.

The procedure used to evaluate this application was the same as in the previous section: selecting types of review materials and activities (enhancing the instructor's proposals with suggestions from the assistant), creating content (five for each category), choosing evaluation criteria (based on proposals from the custom GPT), and assessing the created materials.

Regarding the review materials, here are the types, evaluation criteria, and overall ratings:

1. Summaries. Criteria: clarity and conciseness, organization, and relevance. Score: 4.

2. Outlines. Criteria: structure, comprehensiveness, and readability. Score: 5.

3. Highlights. Criteria: information selection, brevity, and coverage. Score: 4.33.

4. Chronologies. Criteria: accuracy, relevance, and clarity. Score: 5.

5. Concept maps. Criteria: clarity of relationships, thematic coverage, and visual organization. Score: 5.

6. Comparative lists. Criteria: pertinence of compared elements, clarity and organization, and balance and comprehensiveness. Score: 5.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

The course assistant proved to be a valuable tool for studying by creating review materials. It generated high-quality content aimed at enhancing comprehension and learning, with well-written, organized, and structured outputs. It effectively selected relevant information and appropriately used external sources beyond the specific knowledge base to complement its responses. The assistant required detailed instructions only when generating summaries and highlights.

For activities, unified correction criteria were applied to evaluate all 12 types: variety of topics and difficulty, accuracy of questions, and effectiveness of corrections. Below are the 12 categories, along with their average evaluation scores:

1. Multiple-choice questions: 4.33.
2. True/false activities: 4.67.
3. Fill-in-the-blank activities: 3.33.
4. Short-answer questions: 5.
5. Long-answer questions: 5.
6. Matching questions: 4.67.
7. Comparison questions: 5.
8. Historical arguments: 5.
9. Error localization activities: 4.
10. Fictional debates: 5.
11. Decision-making games: 4.67.
12. Role-playing dynamics: 5.

The average rating for the activities was 4.64 out of a maximum of 5, demonstrating that the assistant can be a highly effective tool for promoting learning and enabling students to self-assess their knowledge. As shown, the assistant can generate a wide range of activities, allowing it to address various pedagogical objectives and learning styles.

Activities such as fictional debates, decision-making games, and role-playing dynamics stand out for their ability to engage students and encourage interaction with the assistant, fostering exploration of diverse perspectives. Argumentation exercises, error localization, matching, comparison activities, and long-answer questions stimulate critical analysis and reflection. Meanwhile, multiple-choice questions, true/false activities, fill-in-the-blank exercises, and short-answer questions are useful for reviewing basic concepts.

Despite the high evaluation scores, generative artificial intelligence still has weaknesses and areas for improvement. These include inaccuracies in the phrasing of multiple-choice, fill-in-the-blank, and error localization questions, occasional reliance on external sources unrelated to the course content, and a lack of thematic variety in some activity categories.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

With minor adjustments to the prompts, the assistant could become an excellent complement to students' learning processes.

4.8. Use 8: Preparation of the Theoretical Evaluation

The last three uses are related to planning the assessment of the course. The first focus was on preparing for the theoretical exam. In previous years, the exam consisted of two parts: a multiple-choice questionnaire on Pre-Columbian America and a historical argumentation question on Colonial America. The assistant was used to improve the exam structure and design a new model aligned with the learning objectives and competencies outlined in the course syllabus.

The exam is 90 minutes long and has a maximum length of six pages. Ultimately, the multiple-choice questionnaire and the argumentation question were retained, while short-answer questions and a task involving the creation of an outline were added. The final configuration of the exam is as follows:

1. Multiple-choice questionnaire: 2 pages, 20 questions, an estimated 20 minutes, and 25% of the final grade.

2. Short-answer questions: 1 page, 3 questions, an estimated 15 minutes, and 15% of the final grade.

3. Outline: 1 page, 1 question, an estimated 20 minutes, and 25% of the final grade.

4. Essay question: 2 pages, 1 question, an estimated 35 minutes, and 35% of the final grade.

Once the exam structure was finalized, the assistant was tasked with generating a set of questions for each category to assess its efficiency in this task. Regarding the multiple-choice questions, the assistant created 80 questions. The number of questions generated for each topic depended on its relative length in pages. For example, Topic 1 spans 76 pages, accounting for 30% of the total 253 pages, so 24 questions were generated (30% of 80).

In addition to the number of questions per topic, the GPT assistant was provided with specific requirements:

- Validity of the questions: All questions had to include four options, with only one correct answer. Questions could not have self-evident correct answers or multiple valid responses.

- Difficulty level of the questions: The assistant was instructed to make 25% of the questions as easy, 50% as medium, and 25% as difficult.

- Exclusive use of course content: All questions had to be derived solely from the lecture notes covering the eight topics of the course, which were included in the custom GPT's knowledge base. No other training sources could be used.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Before proceeding with the generation of multiple-choice questions for each topic, the assistant was informed that its performance would be evaluated based on its adherence to these three requirements.

The assistant began generating multiple-choice questions, and the evaluation was very positive, as it adhered to the instructions regarding the number, structure, and distribution of difficulty levels. Of the 80 questions created, 94% were valid, 73% had correctly assigned difficulty levels, and 94% were derived from the topics manually uploaded to the assistant. Additionally, the GPT excelled in speed, efficiency, and the significant time savings achieved by creating a question bank in this manner.

For the short-answer questions, the assistant supported the specification of six types: explanation of key concepts, identification of causes or consequences, brief comparisons, relationships between events and processes, impact evaluations, and analyses of historical perspectives. The assistant was asked to generate eight questions for each category (two easy, four medium, and two difficult). It was informed that the questions would be evaluated based on their validity (clarity, accuracy of wording, and thematic relevance), correct assignment of difficulty level, and exclusive reference to content from the course lecture notes. The results were once again excellent. 90% of the questions met the quality requirement, 77% had correctly assigned difficulty levels, and 96% referred exclusively to the course lecture notes. The only noted weakness was that questions created across different categories tended to be somewhat repetitive. This issue could be addressed by specifying thematic variety as a requirement in the generation prompt.

For the outline question, six types were specified: chronological, comparative, causal, hierarchical, process-oriented, and thematic. The assistant was tasked with creating a total of 24 questions, meeting two main requirements: four questions per category and three questions for each of the eight topics. Regarding difficulty distribution, each category needed one easy question, two medium-difficulty questions, and one difficult question. The evaluation used the same criteria: question validity (clarity, precision, and relevance), correct assignment of difficulty level, and exclusive reference to the course lecture notes. The results were also positive. 83% of the questions exceeded the required quality standard, 88% had the difficulty level correctly assigned, and 96% referred to content from the lecture notes. As an area for improvement, it was identified that the quality of the questions could be enhanced if they addressed more specific or concrete aspects of the topics, which could likely be resolved by providing more detailed instructions in the prompt.

For the essay question, six types were also specified: argumentation, relationships, comparisons, impact evaluations, critical reflections, and process explanations. The assistant was asked to create 24 questions, adhering to the following requirements: four questions per category and three questions for each of the eight topics. Each category needed one easy question, two medium-difficulty questions, and one difficult question. Additionally, the assistant was instructed that the generated questions should allow students to write up to two handwritten pages in the exam. Once the GPT-generated questions were evaluated, 88% were deemed valid, 83% had correctly assigned difficulty levels, and 100% referred exclusively to the content of the lecture notes. However, some questions were overly specific, narrowly meeting the requirement of allowing two pages of written content.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

Overall, the assessment of question creation can be considered nearly outstanding. The assistant could improve (likely with more precise prompts) in assigning difficulty levels and creating more pertinent and varied questions. Nevertheless, its use represents a significant advancement in designing the new exam, specifying question types, and generating a large bank of valid questions for evaluation purposes.

4.9. Use 9: Support in Evaluating Assignments

After creating the practical activities and the theoretical exam, the next step was to assess how the assistant could aid in evaluation by proposing assessment criteria and creating rubrics to streamline grading.

For the daily activities that evaluated classroom participation, a basic and straightforward set of criteria was necessary due to the large number of submissions and their relatively low impact on the final grade (10%). With 97 enrolled students working in pairs, approximately 600 assignments had to be graded (50 pairs across 12 sessions). The accepted evaluation criteria were historical accuracy and depth (60% of the final score), structure, writing, and presentation (20%), and use of the assistant (20%).

For the course's main practical assignment (creating a prompt to interact with a historical figure) a more detailed set of criteria was specified, given its higher weight in the final grade (40%). The chosen criteria were historical accuracy (30% of the final score), prompt design (15%), results (20%), critical analysis and reflection (25%), and quality of the report (10%).

The course assistant correctly and thoroughly defined the evaluation criteria for both types of assignments, adapting them to the specific needs of each and to the requirements outlined by the instructor. Only minor adjustments to the weightings were necessary, primarily to emphasize elements such as historical accuracy and critical thinking. In both cases, the assistant generated complete and explicit rubrics, which greatly facilitated the grading process.

4.10. Use 10: Support in Exam Evaluation

The assistant also contributed to defining evaluation criteria for the different types of exam questions (short-answer questions, outline questions, and essay questions) and to creating rubrics. The criteria and their respective weightings for each type of question were as follows:

- Short-answer questions: Pertinence and accuracy (35%), structure and clarity (10%), contextualization (10%), depth of analysis (15%), use of evidence and specific data (10%), and synthesis skills (20%).

- Outline question: Pertinence and accuracy (35%), organization and structure (15%), synthesis skills (15%), relationships between concepts (10%), contextualization (10%), and thematic coverage (15%).

- Essay question: Pertinence and accuracy (30%), depth (15%), critical reasoning (15%), organization and coherence (10%), contextualization (10%), use of evidence and examples (10%), and writing and style (10%).

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

The assistant quickly and efficiently created both the proposed criteria and the evaluation rubrics for the three types of exam questions. Only minor adjustments to the weightings of the criteria were necessary.

5. Conclusions and Discussion

5.1. Summary of Study Findings

The analysis of the ten uses of the custom GPT assistant highlighted both the strengths and limitations of generative artificial intelligence in teaching planning and the preparation of didactic content for the course America: History from Colonization to the Present.

In the review of the course syllabus (Use 1), the assistant demonstrated a notable ability to improve existing texts and, to an even greater extent, to create new texts related to the course description, objectives, and evaluation criteria. However, its performance in generating curated lists of relevant links for the subject matter was poor.

In the creation of practical activities (Use 2), the assistant proved helpful in generating recommendations and proposals for activities and projects. It also excelled in the development of associated documentation, such as student guides and presentation materials.

The assistant's greatest success was its support in the creation of new lecture notes for the course (Use 3). Over the course of ten days, it helped the instructor produce 253 pages of lecture notes, significantly saving time and greatly improving the quality of the original materials. The custom GPT excelled in tasks such as developing outlines and thematic structures, critiquing previous versions, and drafting content with clear and relevant style. However, certain deficiencies and areas for improvement were identified. These included a lack of depth and examples in its recommendations and difficulties in fully utilizing the instructor-provided sources in the assistant's knowledge base. Additionally, while the canvas system was useful for editing and integrating human and AI contributions, occasional technical issues arose in the texts, such as repetitive content and inconsistencies in information organization.

In the evaluation of lecture notes (Use 4), the assistant excelled in creating the analysis procedure and selecting evaluation criteria, providing high-quality feedback and improvement suggestions. However, the reviews were not always comprehensive and lacked the necessary depth to address complex aspects, limiting their overall effectiveness.

The bibliography proposal (Use 5) was, by far, the assistant's weakest contribution. Although the custom GPT was able to recommend references, these were predominantly focused on English-language texts, often neglecting Spanish-language scholarship, which is essential in the field of Pre-Columbian and Colonial Hispanic American history. Furthermore, the recommendations lacked sufficient reliability. As a result, the instructor ultimately performed this task manually, consulting specialized academic search engines.

In contrast, the proposal and creation of supplementary materials (Use 6) demonstrated the assistant's strong capabilities. It was particularly effective in proposing and generating

diverse, high-quality materials and in establishing evaluation criteria for them. This use stood out for its speed, focus, and creativity.

The assistant also performed exceptionally well in the autonomous student training module (Use 7). It was helpful in recommending training formats and creating review materials and self-assessment activities with high validity, showcasing its potential to support students' independent learning.

In the area of theoretical evaluation (Use 8), the assistant helped the instructor redefine the configuration of the course's final exam and was highly efficient in creating question banks for various types of questions.

In the final two uses, related to support in evaluating assignments (Use 9) and exams (Use 10), the GPT assistant also demonstrated its utility. In both cases, it contributed to the definition of evaluation criteria and the creation of grading rubrics, standing out for the quality and pragmatism of its recommendations.

Overall, the results indicate that the custom GPT assistant is a valuable tool for optimizing teaching planning and content preparation processes. However, its limitations in tasks such as source recommendation and occasional inefficiency and lack of depth in certain contributions highlight the need for constant supervision and technical adjustments to fully realize its potential.

5.2. Practical Implications

The findings of this study highlight the versatility of the custom GPT assistant as a valuable tool for instructors in various tasks related to the planning and preparation of History courses. Its ability to create and enhance content, as well as to develop procedures, stands out as one of its greatest strengths. This flexibility allows the assistant to adapt to the specific needs and requests of instructors, optimizing the design of educational materials and pedagogical processes.

One of the primary advantages of using the GPT assistant is the significant time savings it offers to educators. By taking on repetitive or low-intellectual-demand tasks, the assistant reduces the workload, enabling instructors to focus their efforts on more complex and strategic activities, such as personalizing materials and engaging directly with students. Additionally, the assistant contributes to improving the quality of instructors' work, both in methodological terms and in the creation of educational content.

However, the study also identifies technical areas in need of improvement. These include limitations in bibliography recommendations, the optimization of ChatGPT's canvas system, and the need to enhance the assistant's ability to effectively utilize uploaded materials in its knowledge base. Moreover, it would be crucial to specify the assistant's proprietary training sources to ensure greater transparency and reliability.

Another critical aspect identified is that the quality of the responses generated by the assistant directly depends on the adequacy of the prompts provided by the instructor. This underscores the need for initial training in designing effective prompts to maximize the tool's potential. Additionally, human verification and validation of the AI-generated responses are essential to ensure their accuracy and relevance.

ISSNe: 2386-8864
DOI: 10.6018/pantarei/653731

The study confirms that the GPT assistant should be viewed as a complementary support tool, not a replacement for the instructor. Its role can be fundamental in optimizing processes, but always under the teacher's supervision. Furthermore, the effective use of the assistant's capabilities not only facilitates teaching planning and the preparation of educational materials but also significantly enhances the instructor's knowledge of the subject, both in terms of content and pedagogy. This reinforces the idea that the integration of GPT assistants should become a strategic resource for improving university teaching.

5.3. Relationship with Previous Studies

The body of scientific literature on generative artificial intelligence in education has grown significantly in the past two years. Most published studies focus on enumerating the possibilities and risks of these tools, highlighting their potential to impact teaching and learning. However, studies exploring their practical application in specific educational contexts remain scarce, with even fewer addressing their use in humanities disciplines like History.

This study aims to help fill this gap by focusing on underexplored tasks, such as the use of generative AI in teaching planning and the preparation of specific content for a university course. This practical and discipline-specific approach offers a novel perspective, applying a custom GPT assistant to address concrete challenges in teaching Pre-Columbian and Colonial Hispanic American History.

The findings of this study confirm many of the capabilities highlighted in scientific literature, particularly in terms of time savings, flexibility, and support in content generation. However, they also reinforce one of the most common warnings in previous research: the necessity of human supervision to ensure the accuracy, relevance, and quality of materials created by artificial intelligence. This reliance supports the idea that generative AI tools should be seen as complements that enhance teachers' capabilities, rather than as replacements. In this sense, the study not only validates the general observations found in existing literature but also provides empirical evidence of their application in a specific educational setting.

5.4. Epistemological and Historiographical Considerations

Beyond its technical and pedagogical dimensions, the use of generative artificial intelligence in History education raises relevant epistemological and historiographical questions. Large language models are trained on extensive textual corpora that reflect dominant academic canons, prevailing narratives, and unequal patterns of representation. As a result, their outputs may tend to reproduce conventional interpretations or reinforce established historiographical frameworks unless their use is carefully mediated.

The findings of this study provide empirical support for these concerns. Although the custom GPT assistant generally produced coherent and usable historical outputs, the analysis identified a limited but noticeable tendency toward eurocentric perspectives in certain topics, together with significant weaknesses in bibliographic recommendations—particularly regarding Spanish-language scholarship. These issues illustrate how generative AI tools may inadvertently privilege dominant academic traditions and

linguistic contexts, thereby constraining the plurality of historical narratives available to students.

At the same time, the results indicate that such epistemological risks are neither inherent nor unavoidable. When embedded within a clearly defined pedagogical and historiographical framework, the assistant's outputs can be effectively shaped through explicit prompting strategies, the use of instructor-curated and verified sources, and systematic human review. In this respect, prompts that explicitly required the inclusion of diverse perspectives, the consideration of historiographical debates, or the visibility of historically marginalized actors proved effective in counterbalancing default tendencies in the model's responses.

From a didactic standpoint, these findings reaffirm the central role of the instructor as an epistemic mediator. Generative AI tools do not exercise historical judgment; rather, they tend to amplify existing assumptions unless guided by explicit epistemological criteria. Their educational value therefore depends not only on technical performance, but on the clarity of the historiographical orientation, the quality and diversity of the sources employed, and the instructor's capacity to integrate AI-generated content into a reflective and critical teaching practice. In this sense, the responsible use of generative artificial intelligence in History education requires treating such tools not as neutral conveyors of knowledge, but as instruments whose outputs must be continuously interrogated, contextualized, and revised in accordance with the discipline's epistemological standards.

5.5. Limitations and Opportunities for Future Research

While this study provides detailed insight into the integration of a custom GPT assistant in teaching planning and content preparation for a university-level History course, its findings should be interpreted within the boundaries of its specific context. The research was conducted in a fourth-year History course at the University of Alicante, which allows for well-grounded conclusions regarding its application in this setting but limits the immediate generalizability of the results to other disciplines, educational levels, or institutional environments. Future research should therefore examine the replicability of these findings across different academic contexts and curricular structures.

One of the most significant limitations identified concerns the reliability and control of bibliographic recommendations generated by the assistant. The empirical analysis revealed a very high rate of fabricated or unverifiable references in Spanish-language outputs (approximately 90%), compared to a substantially higher proportion of verifiable references in English (around 85%). Consequently, all bibliographic suggestions required systematic verification by the instructor using specialized academic databases. This finding underscores the need to treat AI-generated bibliographic outputs as provisional and highlights the importance of explicit source-verification protocols when integrating generative AI into academic teaching and research practices.

The study also identified the presence of implicit biases in some of the assistant's outputs, particularly a tendency toward eurocentric perspectives in certain historical topics. Although these biases were generally limited in scope and could be mitigated through careful prompt design and the use of instructor-curated knowledge bases, their occurrence reinforces the necessity of continuous human supervision. In History education—where

narrative construction, historiographical balance, and the representation of historically marginalized actors are central concerns—the uncritical use of generative AI poses clear epistemological risks. Addressing these risks requires deliberate pedagogical strategies, including counter-bias prompting, diversified and verified bibliographic inputs, and systematic critical review of generated content.

From a methodological standpoint, the evaluation of the assistant's outputs was conducted by a single rater—the instructor responsible for the course—which introduces the possibility of single-rater bias. This limitation was partially mitigated through iterative verification processes, including the systematic comparison of AI-generated outputs with course objectives, historiographical standards, and pre-existing teaching materials. Nevertheless, future studies could strengthen methodological rigor by incorporating additional forms of triangulation, such as peer validation of selected outputs or structured student feedback, to provide complementary perspectives on the assistant's effectiveness and limitations.

The effective use of the custom GPT assistant was also found to depend heavily on the instructor's experience in prompt design and in supervising AI-generated content. This reliance highlights the importance of targeted professional development and opens a line of future research focused on identifying training strategies that facilitate the responsible and effective adoption of generative AI across different levels of teaching experience and digital competence.

From an ethical and institutional perspective, the study aligns with established principles regarding the responsible use of artificial intelligence in higher education. In accordance with institutional guidelines at the University of Alicante (2025), the assistant was conceived as a support tool rather than a substitute for academic judgment, with explicit emphasis on transparency, human oversight, and the mitigation of potential biases. At a broader level, the design and use of the assistant are consistent with emerging European regulatory frameworks, such as the EU Artificial Intelligence Act (European Union, 2024), which emphasize a risk-based approach, accountability, and the central role of human supervision in the deployment of AI systems. Although this study does not constitute a regulatory analysis, these principles provide a relevant ethical framework for interpreting its findings.

Future research could build on this work by expanding the research team to include specialists in didactics, digital pedagogy, and educational artificial intelligence, thereby enabling more interdisciplinary analyses. In addition, empirical studies focusing on student learning outcomes, motivation, and the development of critical digital competencies would contribute to a more comprehensive understanding of the educational impact of generative AI tools.

Despite these limitations, this study provides empirical evidence of the concrete ways in which a custom GPT assistant can support teaching planning and content preparation in university-level History education. By systematically analyzing ten instructional uses within a real teaching context, the research offers a use-oriented and discipline-specific perspective that remains underrepresented in the literature on generative artificial intelligence in education. The findings highlight both the potential and the constraints of these tools, underscoring the importance of clear pedagogical objectives, careful prompt

design, and continuous human supervision. Ultimately, the study reinforces the view of generative AI not as a substitute for the instructor, but as a mediated support resource whose educational value depends on informed, critical, and responsible teaching practice.

Acknowledgments and Funding

This study has been conducted within the framework of the project "Red IA-UA," part of the ICE Network Program for Research in University Teaching, funded by the University of Alicante (Spain).

Specific Contribution of the Authors

Tasks performed by the author of the study: conceptualization, methodology, validation, formal analysis, investigation, resources, writing – original draft, writing – review & editing, project administration, funding acquisition.

Bibliography

Acun, C., & Acun, R. (2023). GAI-Enhanced Assignment Framework: A Case Study on Generative AI Powered History Education. In *NeurIPS'23 Workshop on Generative AI for Education (GAIED)*.

Almasri, F. (2024). Exploring the Impact of Artificial Intelligence in Teaching and Learning of Science: A Systematic Review of Empirical Research. *Research in Science Education*, *54*, 977–997. https://doi.org/10.1007/s11165-024-10176-3

Ayeni, O., Al Hamad, N., Chisom, O., Osawaru, B., & Adewusi, O. E. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, *18*(2), 261-271. https://doi.org/10.30574/gscarr.2024.18.2.0062

Bertram, C., Weiss, Z., Zachrich, L., & Ziai, R. (2021). Artificial intelligence in history education. Linguistic content and complexity analyses of student writings in the CAHisT project (Computational assessment of historical thinking). *Computers and Education: Artificial Intelligence*, 100038. https://doi.org/10.1016/j.caeai.2021.100038

Calderón Pantoja, N. (2024). Hacia una inclusión de la Inteligencia Artificial en la enseñanza de la Historia: un acercamiento a la ingeniería de prompts para el trabajo docente. *Reseñas de enseñanza de la Historia*, *25*, 83-99. https://revele.uncoma.edu.ar/index.php/resenas/article/view/5571

Carrasco-Rodríguez, A. (2023). Reinventando la enseñanza de la Historia Moderna en Secundaria: la utilización de ChatGPT para potenciar el aprendizaje y la innovación docente. *Studia Historica: Historia Moderna, 45*(1), 101-145. https://doi.org/10.14201/shhmo2023451101146

Carrasco-Rodríguez, A. (2024a). Evaluation of Artificial Intelligence tools (ChatGPT and Google Bard) in higher education for the history of America: answering questions and crafting assignments with a gender perspective. In M. Ibáñez & A. Vega (coord.), *Building knowledge: visions from education and the humanities* (pp. 49-68). Peter Lang. https://doi.org/10.3726/b21877

Carrasco-Rodríguez, A. (2024b). Perceptions of Generative Artificial Intelligence Among University Early Modern History Students. *Tiempos Modernos: Revista electrónica de*

Historia Moderna, *14*(49), 269-285. http://www.tiemposmodernos.org/tm3/index.php/tm/article/view/5927

Carrasco-Rodríguez, A., Navarro-Colorado, B., Zurita-Aldeguer, R., Torregrosa-Peinado, H., López-Pinel, M., & Pérez-Llorca, J. (2024). Game-Based Learning, Synergies with Artificial Intelligence, and Educational Innovation in University Teaching of History and Multimedia Engineering. In R. Satorre-Cuerda (coord.), *Redes de Investigación e Innovación en Docencia Universitaria* (pp. 47-62). Instituto de Ciencias de la Educación, Universidad de Alicante. http://hdl.handle.net/10045/149202

Carretero, M., & Gartner, E. (2024). Artificial Intelligence and historical thinking: a dialogic exploration of ChatGPT. *Studies in Psychology*, *45*(1), 80-102. https://doi.org/10.1177/02109395241241379

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The Use of Artificial Intelligence (AI) in Online Learning and Distance Education Processes: A Systematic Review of Empirical Studies. *Applied Sciences*, *13*(5), 3056. https://doi.org/10.3390/app13053056

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Fareed, M., Bou Nassif, A., & Nofal, E. (2024). Exploring the Potentials of Artificial Intelligence Image Generators for Educating the History of Architecture. *Heritage*, *7*(3), 1727-1753. https://doi.org/10.3390/heritage7030081

Fernández-Arrillaga, I., Carrasco-Rodríguez, A., Ávila-Martínez, M., San Mauro-Martínez, I., Beltrán-Pastor, S., & Luz-Fernández, N. (2024). Teaching Strategies for Developing Critical Thinking in University History Students Based on Gender Perspective and Generative Artificial Intelligence. En R. Satorre-Cuerda (coord.), *La docencia universitaria en tiempos de IA* (pp. 38-49). Octaedro. http://hdl.handle.net/10045/149121

García-Peñalvo, F. J. (2024). Inteligencia artificial generativa y educación: Un análisis desde múltiples perspectivas. *Education in the Knowledge Society (EKS)*, *25*, e31942. https://doi.org/10.14201/eks.31942

Garrido-Merchán, E. C., Arroyo-Barrigüete, J. L., Borrás-Pala, F., Escobar-Torres, L., de Ibarreta, C. M., Ortiz-Lozano, J. M., & Rua-Vieites, A. (2023). Real Customization or Just Marketing: Are Customized Versions of Chat GPT Useful? *arXiv*, 2312.03728. https://doi.org/10.48550/arXiv.2312.03728

Gökçearslan, S., Tosun, C., & Erdemir, Z. G. (2024). Benefits, challenges, and methods of artificial intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education*, *7*(1), 19-39. https://doi.org/10.46328/ijte.600

Hutson, J. (2024). Integrating art and AI: Evaluating the educational impact of AI tools in digital art history learning. *Forum for Art Studies*, 1(1), 393.

Idroes, G. M., Noviandy, T. R., Maulana, A., Irvanizam, I., Jalil, Z., Lensoni, L., . . . Idroes, R. (2023). Student Perspectives on the Role of Artificial Intelligence in Education: A Survey-Based Analysis. *Journal of Educational Management and Learning*, *1*(1), 8-15. https://doi.org/10.60084/jeml.v1i1.58

Kindenberg, B. (2024). ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis. *Education Sciences*, *14*(5), 530. https://doi.org/10.3390/educsci14050530

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Lazareva, A., Vindbo, S., & Spanos, A. (2024). Student Experiences with Using ChatGPT in History Classes. In *INTED2004 Proceedings* (pp. 2335-2342). IATED. https://doi.org/10.21125/inted.2024.064

McGrath, C., Pargman, T. C., Juth, N., & Palmgren, P. J. (2023). University teachers' perceptions of responsibility and artificial intelligence in higher education-An experimental philosophical study. *Computers and Education: Artificial Intelligence*, *4*, 100139. https://doi.org/10.1016/j.caeai.2023.100139

Melero-Muñoz, I. (2023). La inteligencia artificial como elemento de innovación: posibilidades, límites y desafíos en la docencia universitaria de la historia. In C. Hervás-Gómez (coord.), *Conexiones digitales: las tecnologías como puentes de aprendizaje* (pp. 550-570). Dykinson.

Mottl, J., & Musílek, M. (2024). Information Technology as a Tool for Teaching History with a Focus on Artificial Intelligence and Audiovisual Material. In *2024 47th MIPRO ICT and Electronics Convention (MIPRO)* (pp. 278-282). IEEE. https://doi.org/10.1109/MIPRO60963.2024.10569216

Onesi-Ozigagun, O., Ololade, Y. J., Eyo-Udo, N. L., & Ogundipe, D. O. (2024). Revolutionizing education through AI: a comprehensive review of enhancing learning experiences. *International Journal of Applied Research in Social Sciences*, *6*(4), 589-607. https://doi.org/10.51594/ijarss.v6i4.1011

Rosser-Limiñana, P., & Soler-Ortiz, S. (2024). Emotions in Digital Learning: Assessing the Emotional and Pedagogical Impact of Historical Simulations through ChatGPT in University Classrooms. In M. D. Díaz-Noguera (coord.), *Artificial Intelligence and Education* (pp. 175-188). Octaedro. https://doi.org/10.36006/09643-1

Sánchez, O. V. (2023). Uso y percepción de ChatGPT en la educación superior. *Revista de Investigación en Tecnologías de la Información*, *11*(23), 98-107. https://doi.org/10.36825/RITI.11.23.009

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, *3*(1), 110-121. https://dx.doi.org/10.2139/ssrn.4378735

Soler-Ortiz, S., & Rosser-Limiñana, P. (2024a). Desafiando los límites del aprendizaje histórico: una propuesta educativa innovadora basada en la pedagogía crítica, IA y ChatGPT para comprender la Guerra Civil española, la dictadura franquista y la transición democrática. In B. Pizà-Mir (coord.), *Las ciencias sociales, las humanidades y sus expresiones artísticas y culturales: una tríada indisoluble desde un enfoque educativo* (pp. 263-283). Dykinson.

Soler-Ortiz, S., & Rosser-Limiñana, P. (2024b). Reescribiendo la historia a través de un Scape Room Juana I, la Reina Cuerda como herramienta educativa innovadora para la sensibilización sobre la violencia de género. In O. Buzón García (coord.), *Aprendizaje 4.0: inteligencia artificial, redes sociales y rol docente en la era digital* (pp. 509-526). Dykinson.

Sperling, K., Stenberg, C. J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in Teacher Education: A scoping review. *Computers and Education Open*, *6*, 100169. https://doi.org/10.1016/j.caeo.2024.100169

Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P., & Cózar-Gutiérrez, R. (2023). From Human to Machine: Investigating the Effectiveness of the Conversational AI ChatGPT in Historical Thinking. *Education Sciences*, *13*(8), 803. https://doi.org/10.3390/educsci13080803

University of Alicante. (2025). *Código de buenas prácticas en investigación*. https://web.ua.es/es/vr-investigacio/documentos/codigo-de-buenas-practicas-en-investigacion.pdf