GenAI Embedded in Activities of Academic Research: Experiences and Lessons from HCI Studies

Tatiana Alencar[1], Leonardo Ciscon[1], Gessé Pereira[1], Maylon Macedo[1], Suéllen Martinelli[1], Joelma Choma[1], Luciana Zaina[1]

[1]Federal University of São Carlos (UFSCar), São Paulo, Brazil

{tatiana.alencar,leonardo.ciscon,

gesse.pereira}@estudante.ufscar.br,

{macedomaylon, suellen.r.martinelli}@gmail.com,

jh.choma@hotmail.com, lzaina@ufscar.br

Abstract. Introduction: *Generative Artificial Intelligence (GenAI) models, particularly ChatGPT, are increasingly being used to support various stages of scientific research.* ***Objective:*** *This paper aims to report on nine specific instances of the author team's experience using ChatGPT to support their research activities within the field of Human-Computer Interaction (HCI). The goal is to identify and present lessons learned from these applications.* ***Methodology:*** *The methodology involved documenting and analyzing nine distinct cases where the author team utilized ChatGPT to assist with different research tasks. The analysis focused on identifying patterns of effective use and drawing conclusions regarding the tool's strengths and limitations.* ***Results:*** *The analysis of these nine cases yielded six key lessons, demonstrating that ChatGPT provides substantial support across several research tasks. These tasks include content refinement, text reduction, insight generation, screening, summarization, and data visualization. The primary benefits observed were accelerated workflows, improved clarity, and enhanced systematic organization. However, the study also revealed that the effectiveness of ChatGPT is highly dependent on the use of precise, well-structured prompts and the segmentation of complex tasks into smaller, manageable units.*

Keywords *GenAI, ChatGPT, UX Research.*

1. Introduction

Generative Artificial Intelligence (GenAI) models have recently gained immense popularity and have been widely applied in various scientific research domains as essential productivity tools [Khalifa e Albadawy 2024]. GenAI models such as Large Language Models (LLMs) are now widely used for academic tasks ranging from the generation of ideas to qualitative data analysis through natural language processing [Rahman et al. 2023]. LLMs like ChatGPT[1] are being increasingly leveraged to assist in literature reviews, generate source code, and support academic writing [Hosseini e Horbach 2023]. As researchers continue to explore GenAI's potential, its integration into academic workflows is reshaping how scientific inquiry and collaboration unfold [Tenhundfeld 2023] [Watkins 2024].

[1]https://openai.com/index/chatgpt/

GenAI-based models and tools offer significant benefits in academic writing and research, enhancing productivity by accelerating text generation, improving grammar and clarity, assisting with structure and flow, and supporting multiple stages of the research process, including brainstorming, literature review, summarization, and translation. However, several challenges and risks need to be addressed, according to [Khalifa e Albadawy 2024]. The use of GenAI can produce misleading or inaccurate content, including incorrect references, which poses a risk to scientific integrity [Wen e Wang 2023]. Ethical concerns around authorship, plagiarism, and over-reliance on AI raise questions about academic responsibility [Zaina et al. 2024]. Constant human oversight is essential, as models cannot reliably interpret figures, tables, or nuanced scientific arguments [Petersen e Gerken 2025]. While powerful, these tools should be treated strictly as assistants, not substitutes for critical thinking and scholarly expertise [Freire et al. 2023].

In this study, we report our experience of using GenAI (i.e., ChatGPT) as a research partner in Human-Computer Interaction (HCI) research, specifically for qualitative analyses, textual reviews, and text coding. This work emphasized the generation of HCI research insights, rather than generic artifacts (e.g., personas) or design processes. The authors' team brings together experience adopting HCI concepts and methods in the software industry, teaching, and university extension. The article explores several applications of LLMs in the context of HCI research, using a case-based approach. The cases present the parties involved, the context of the reported experience, the systematic method applied (i.e., prompts), the contributions of the approach practiced, and reflections on the researchers' experience.

The main contribution of this article lies in the demonstration of concrete practices for the use of GenAI in the context of HCI research, highlighting how these practices can contribute to increased productivity and enhanced quality of research processes. The structure of the article includes, in Section 2, the presentation of the theoretical background; in Section 3, the description of the adopted method; in Section 4, a detailed exposition of nine case studies; followed by the systematization of the lessons learned in Section 5; and, finally, the concluding remarks in Section 6.

2. Background

2.1. GenAI and LLMs

A large language model (LLM) such as OpenAI's ChatGPT is a generative AI (GenAI) model trained on vast amounts of text, enabling it to understand, generate, and interact with human language, spoken and written [Chang et al. 2024]. These models possess a transformation capability by predicting the probability of word sequences and generating coherent text based on a specific input [Brown et al. 2020], known as a prompt. A **prompt** is the instruction or information given to the model to guide its response generation [Schmidt et al. 2024]. The prompt constitutes the main interface between humans and LLMs, acting as the basis through which users give instructions, ask questions, and shape model responses [Chen et al. 2024]. A well-elaborated prompt often leads to a better response. An essential practice for enhancing the performance and reliability of interactions with LLMs is called Prompt Engineering. This practice involves strategically designing and refining prompts to achieve more accurate, relevant, or creative **outputs** from the model [Liu e Chilton 2022].

LLMs process language by breaking down text into smaller units called **tokens**, which can represent words, sub-words, or even individual characters. These tokens are essential for efficiently interpreting and generating text [Vaswani et al. 2017]. A set of tokens represents a **context** that refers to what the model can consider at any given moment when formulating a response, encompassing both the initial prompt and the ongoing conversation. The quality and relevance of an LLM's output often depend significantly on the amount and clarity of the contextual information provided through these tokens [Dong et al. 2024].

Internal **parameters** that serve as weights and biases within neural networks shape a model's understanding of language and ability to generate text [Dong et al. 2024]. These parameters represent the values that the model adjusts during training to learn linguistic patterns, relationships, and knowledge from the data. The number of parameters in a model determines its capability level to represent complex structures and nuances of human language. Temperature is an important parameter that influences the creativity of a model's responses, significantly affecting its behavior. Lower temperature values (e.g., 0.2) produce more focused and deterministic answers, while higher values (e.g., 0.8 or above) produce more diverse, creative, and sometimes less predictable **outputs** [Gilardi et al. 2023].

2.2. GenAI-assisted research

Several studies have explored the impact of GenAI on academic research and its transformation of the research process [Hanafi et al. 2025, Hosseini e Horbach 2023, Lund et al. 2023, Rahman et al. 2023]. [Hanafi et al. 2025] investigated applications of GenAI, especially models such as ChatGPT, at eight stages of the academic research lifecycle in the Engineering and Educational Technology area. The eight stages of the academic research lifecycle were idea generation, literature review, research design, data collection, data analysis, interpretation of results, writing and dissemination, and peer review and publication. In addition, they critically discuss the ethical and practical risks and challenges associated with using GenAI in research, including plagiarism, algorithmic bias, data privacy, cognitive dependency, and the erosion of critical thinking.

[Hosseini e Horbach 2023] analyzed the potential impacts of LLMs applied in the academic peer review process in the areas of Information Science and Scientific Communication. By discussing the benefits, risks, and ethical implications of using AI in this context, the authors provide practical recommendations for responsible use such as training on LLMs in ethics and best practices courses; mandatory disclosure of AI use in editorial reviews or decisions; ensuring data protection when using AI with unpublished manuscripts; and creating clear policies by publishers and post-publication platforms. [Lund et al. 2023] also discusses the impact of ChatGPT on academia and scientific publishing in the Information Science and Technology area, with a particular focus on emerging ethical issues and dilemmas related to authorship, originality, and plagiarism. [Rahman et al. 2023] found that ChatGPT works satisfactorily as an academic writing assistant, offering support in tasks such as idea generation, drafting, and text organization in the area of Educational Technology. However, they pointed out that it does not replace the researcher's role in critical tasks such as problem formulation, literature review, analysis, and interpretation of results. Its usefulness depends directly on the quality of the prompts and the continuous human intervention.

Several studies have also reported different approaches to integrating GenAI tools and models into the various phases of the Systematic Literature Review (SLR) process, focusing on automating and accelerating steps such as study screening, data extraction, and synthesis tasks in the areas of Software Engineering, Industrial Engineering, Medicine and Psychiatry, and Economics and Management [Felizardo et al. 2024, Alshami et al. 2023, Fabiano et al. 2024, Burger et al. 2023]. According to these studies, the primary challenges of using AI in systematic reviews include inaccuracy in decision-making, lack of contextual understanding, and dependence on the quality of input data. Furthermore, technical limitations, such as token restrictions and difficulties in processing full-text articles, which affect the fluidity of analysis, along with the absence of formal protocols (e.g., PRISMA-AI), hinder the standardization and regulation of AI use in systematic literature reviews (SLRs).

Other studies highlighted how GenAI tools have been applied and are impacting qualitative and quantitative research methods [Perkins e Roe 2024, Xiao et al. 2023]. [Perkins e Roe 2024] described a case that integrated ChatGPT into the inductive thematic analysis process, comparing its performance with that of a human analyst in the Social Sciences Communications area. In this study, the authors highlight three limitations: the inconsistency of the results when ChatGPT generated variations in the codes and suggested themes from the same set of analyzed data; the dependence on the researcher's skill in formulating the prompt; and the superficiality of cultural and contextual aspects that could enrich the qualitative analysis. They also explored how GenAI tools (e.g., ChatGPT and Claude[2]) can assist in analyzing, visualizing, and interpreting quantitative data, accelerating the analytical process. They noted several technical and ethical limitations, including the risk of obtaining misleading correlations due to excessive testing. Additionally, they noted the lack of transparency in how the AI arrived at its calculations. There is also a risk of misinterpretation if researchers do not critically evaluate the outputs, especially those with little experience in statistics. [Xiao et al. 2023] investigated how ChatGPT-3 [Brown et al. 2020] can support deductive qualitative analysis, where a predefined set of codes (codebook) is used to label qualitative data. In this study in the HCI area, the performance of ChatGPT-3 is compared with the results of expert human coders. As a result, they highlight a reasonable level of accuracy due to more structured tasks (i.e., deductive coding), which did not require training or fine-tuning, making practical application more accessible to researchers with less technical knowledge.

Research on the usage of GenAI in academic activities has been conducted by researchers from various scientific areas, as mentioned above. However, only one study discusses the experiences of researchers utilizing GenAI specifically in HCI research [Xiao et al. 2023]. Additionally, most studies report practical application of GenAI for particular tasks, such as coding for qualitative analysis [Xiao et al. 2023], reviewing academic articles [Hosseini e Horbach 2023], literature review [Alshami et al. 2023, Fabiano et al. 2024, Burger et al. 2023], and quantitative analysis [Perkins e Roe 2024]. While the remaining studies explored the use of GenAI in various stages of research based on studies from the existing literature [Hanafi et al. 2025, Rahman et al. 2023, Lund et al. 2023, Perkins e Roe 2024]. In contrast to the article by [Hanafi et al. 2025],

[2]https://claude.ai/

which classifies the use of GenAI into eight research stages, we considered the three research phases proposed by [Lazar et al. 2017] to categorize the cases addressed in this work. A description of each research stage can be found in Section 3.

3. Method

Regarding the development of the cases, initially, all authors wrote about their experiences with ChatGPT for HCI research on various topics. We wrote the cases in a file shared among all authors without a predefined format. Afterward, the fourth author developed a template to help the other authors standardize the case presentation. This template was also discussed and validated among all authors. Therefore, each author rewrote their own cases based on the template, which specifies some items such as research objective, researcher experience, prompt, and output example. Thus, we structured nine complete cases that follow the same presentation pattern.

From our nine cases, we conducted a qualitative analysis with the aim of extracting themes that would contribute to the elaboration of lessons learned. Our qualitative analysis was guided by two approaches. A deductive approach is useful by present a logic similar to quantitative research (common in Computer Science), being a top-down process where the researchers start the analysis with supported by a list of codes determined beforehand [Gibbs 2018]. However, an inductive approach aims at data-driven coding to generate themes or new findings to interpret the reality or actions of people inserted in a context [Charmaz 2006]. Therefore, our study combined the *closed coding* and *open coding* techniques, which aim at these two approaches [Martinelli et al. 2023]. Both techniques condense textual data into a concise format reflecting its meaning and significance. The individual components of the text are then combined to form a single label (i.e., a code) that symbolizes a concept or idea [Gibbs 2018].

In our first round of qualitative analysis, we began the analysis by closed coding to identify the most suitable categories for each case, based on the purpose of the cases within the context of the HCI research process. We adopted the research phases in HCI proposed by [Lazar et al. 2017] as a reference to categorize our cases. According to the authors, the research process in HCI can be divided into three main phases: (1) designing research, (2) running data collection and analysis, and (3) reporting research. [Lazar et al. 2017] highlights that all scientific disciplines seek excellence in each of these stages, and, for this reason, this general research design model is suitable to guide various studies in the field of HCI. Therefore, we assigned one or more categories (i.e., codes) related to the phases of the HCI research process, regarding the research goal of each case. After, we examined in all cases the items about research goal, outputs, and researcher experiences from open coding. We applied one or more initial codes to these items. Figure 1 illustrates examples of quotes about items of a case and the respective codes assigned to them. The closed codes (with square) and open codes (with triangle) applied appear on the side of the figure.

The first round resulted in seventeen initial codes that returned a set of benefits, weaknesses, and learnings that researchers reported about using ChatGPT in HCI research. All cases were stored in a Google Sheets project and analyzed by two authors (fifth and sixth). Both conducted the qualitative analysis asynchronously, applying closed and open codes in columns specific to each author. The two authors have more than six

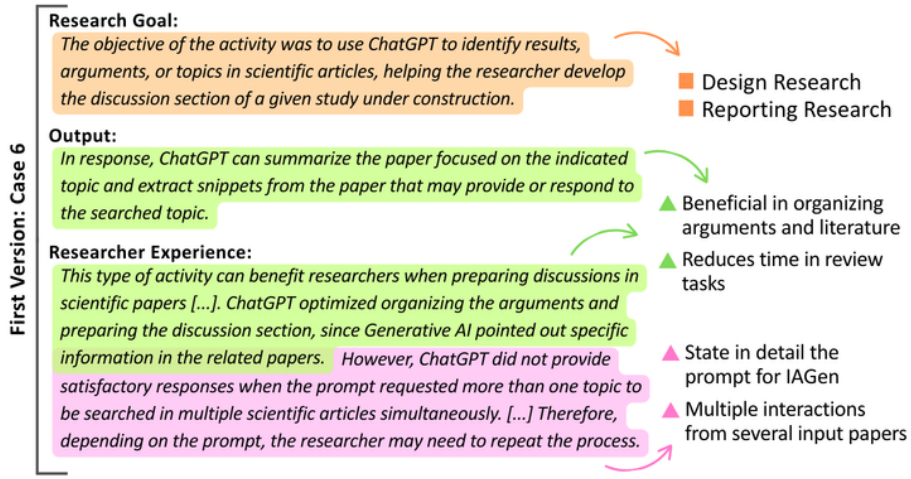

Figura 1. Example of quotes analyzed and codes applied.

years of experience in qualitative analysis in HCI academic research. At the end of the first round, the same authors participated in a consolidation meeting to discuss the codes applied in each case (agree or disagree) and generate initial themes capable of grouping similar codes.

Since different cases presented similar content about goals and outputs, the second round of qualitative analysis began by rewriting the cases. The fourth author was responsible for this activity to minimize bias in rewriting the cases. This author is a Ph.D. candidate with more than six years of experience with HCI academic research and has dedicated the last year to research on GenAI. Thus, the fourth author rewrote the cases so that each experience would report unique goals and outputs, besides considering the codes applied in the qualitative analysis in this new version of the cases. Again, we consolidated nine cases with unique research objectives and specific output examples for each case (Section 4). The fifth and sixth authors reanalyzed these cases in the second round to reapply the closed and open codes identified previously. The authors developed another consolidation meeting to discuss the codes applied and then to reach a consensus on the emerging themes. Thus, six themes were consolidated by common agreement between the two authors.

The six extracted themes and their related codes guided the development of the lessons learned. Therefore, during the third round, the first and sixth authors wrote each lesson learned from one specific theme (Section 5). The first author is a Ph.D. candidate with more than five years of experience in the software industry and has dedicated the last year to using GenAI in their HCI academic research. Both authors had access to the Google Sheets project where the qualitative analysis was performed to consult all the codes and themes. Based on the lessons learned, the third round was concluded with all authors reviewing the cases and lessons. This review also included a final meeting between all authors to discuss adjustments to the case template, adding items such as use context, prompt role, and output explanation. The second and third authors are Ph.D. candidates with experience in the software industry and dedicated the last year to using GenAI in their HCI academic research, while the seventh author is a senior researcher in HCI with more than fifteen years of experience.

4. Cases

Our nine cases are organized into four items that characterize the general information about each case, besides others eight items that describe the specific information. The general information are defined as *research goal* (i.e, a brief description of the purpose of the research activity), *GenAI tools* (i.e., LLM tools used to assist in the activity), *outputs* (i.e., resume about the results obtained from the interaction with GenAI), and *researcher experience* (i.e., the benefits and difficulties perceived by researchers in carrying out tasks assisted by GenAI). In addition, the specific information is represented by the items *use context*, *research objective*, *input sources*, *tools/plataforms*, *prompt role*[3], *prompt*, *output explanation*, and *output example*.

4.1. Case 1 - Text Review

The **research goal** of this case consisted of using ChatGPT as a reviewer of scientific texts written in English, which involved correcting grammar and subject-verb agreement issues. The papers were written in the context of a doctoral project about UX Research, Long-Term UX, and software startups. To assist in the development of this task, the **GenAI tool** utilized was ChatGPT with Write for Me[4]. The **outputs** from the interaction with ChatGPT included a revised version of the paragraph submitted for analysis, presenting corrections in bullet points along with the reason for the change. The **researcher experience** during the execution of this task involved a trade-off between efficiency and careful interaction. Requesting ChatGPT to revise an English paragraph typically returns only the revised version of the text. If the user wishes to understand the rationale behind each correction or receive justifications for the changes made, the prompt must explicitly request this information. To obtain detailed feedback on the revision process, a prompt was developed that requested the ChatGPT to clarify which errors were identified and how each of them should be corrected. Additionally, the researcher sought to work on one paragraph at a time with ChatGPT to minimize the risk of hallucinations from the ChatGPT. A prompt utilizing this strategy and an output with individual corrections (i.e., a part of the paragraph) are shown below.

Prompt case 1: Text review

a) Use Context: Development of a scientific paper on UX Research practices and Long-Term UX cycle.

b) Research Objective: To investigate what UX Research practices software startups apply and how this work is connected with the Long-Term UX cycle.

c) Input Source: A paragraph originally written in English by the author.

d) Tools/Platforms: ChatGPT from Write for Me.

e) Prompt Role: You are a UX researcher with strong academic experience in Human-Computer Interaction (HCI), qualitative research methods, and UX Research practices with a longitudinal perspective (Long-Term UX). Your task is to evaluate each paragraph written in English and suggest grammatical and spelling corrections.

f) Task to perform: Identify grammatical, spelling, and agreement errors in scientific texts. The answer should include explanations for the errors identified and alternatives for correction.

[3]Prompt roles were empirically derived and refined based on output quality and chain-of-thought analysis, not pre-defined models.

[4]Solution dedicated to writing, rewriting, and proofreading of texts. Available in: https://chatgpt.com/g/g-B3hgivKK9-write-for-me.

Prompt case 1 (continuation): Text review

g) Prompt: *Chat, I would like you to review the grammar and subject-verb agreement of the following paragraph. If you make any corrections, please describe what they were:* [paragraph written in English to be corrected].

h) Output Explanation: ChatGPT returned a revised version of the paragraph submitted for analysis, presenting corrections sentence by sentence, besides reason or justification for the correction.

i) Output Example:
Original: "While be-goals refer to users' emotional needs, such as feelings of satisfaction or enthusiasm caused by using the product or service."
Revised to: "while be-goals address users' emotional needs, such as feelings of satisfaction or enthusiasm derived from using the product or service."
Reason: I combined the sentences to avoid fragmentation and improve the flow of ideas. I replaced "caused by" with "derived from," which is more formal and fluent in scientific writing.

4.2. Case 2 - Article Screening

The **research goal** of this case consisted of leveraging GenAI across different literature review methodologies, specifically in a systematic review and a gray literature review in the context of a doctoral project. A systematic review is a structured method of identifying, evaluating, and synthesizing all available research on a particular topic using a rigorous and replicable methodology [Liberati et al. 2009]. A gray literature review involves the examination of non-peer-reviewed sources such as reports, theses, government documents, and other unpublished or informal publications [Lefebvre et al. 2008]. In both approaches, GenAI was used to support the screening process by acting as a second reviewer and applying predefined inclusion and exclusion criteria. To assist in this task, GenAI was accessed through the ChatGPT interface for manual interaction and evaluation of the articles. The **outputs** resulting from the interaction with ChatGPT included classified lists of articles based on defined criteria. A key feature of the output was the inclusion of textual justifications - direct excerpts from the articles - to support each YES or NO classification, which enhanced transparency, traceability, and trust in the screening decisions. The satisfactory Kappa concordance index [Fleiss et al. 2003] of 0.61 was achieved by the model in the context of the academic literature review. The Kappa coefficient represents the level of agreement between evaluators and results in a number between -1 and 1. The **researcher experience** during the execution of this task presents a beneficial increase in efficiency and transparency through the justification of decisions. However, this advantage is tempered by the persistent risk of unintentionally excluding relevant articles, particularly with ambiguous text. As a result, the researcher must process documents in smaller segments to mitigate potential errors and maintain the quality of the screening process. This necessitates a strategic approach to balance the ChatGPT's speed with the need for careful oversight and adaptation of the workflow. The developed prompt is shown below.

Prompt case 2: Article Screening

Example 1: Classification of articles in the systematic review based on [Alshami et al. 2023]

a) Use Context: Systematic review in the field of Human-Computer Interaction and Smart Cities.

b) Research Objective: To support article screening through ChatGPT based on predefined inclusion and exclusion criteria.

c) Input Source: Structured tables summarizing metadata and abstracts of peer-reviewed articles, organized in spreadsheet format (columns A to D).

Prompt case 2 - Example 1 (continuation): Article Screening

d) Tools/Platforms: GPT-4 (OpenAI), accessed through the ChatGPT web interface.

e) Prompt Role: You are a researcher in Human-Computer Interaction and Smart Cities. Your role is to analyze each record and determine whether it describes a co-design initiative involving users and the use of IoT technologies.

f) Task to perform: Read the summarized content (metadata and abstract) of each article and classify it as relevant or not.
g) Prompt: *First, explain in your own words what you understand by: (1) Smart city co-design projects involving user collaboration; (2) Application of technologies associated with the Internet of Things (IoT) in smart city projects. Then, read columns A through D of the attached table and classify each work into: 1. Smart city co-design projects involving user collaboration and IoT technologies; or 2. Not related. Justify your decision based on the provided information.*

h) Output Explanation: ChatGPT returns a classification label for each article (e.g., "1" or "2") along with a short explanation or justification grounded in the content of the summary.

i) Output Example:
YES
Meets the criteria for IoT and smart cities with sufficient detail.

Prompt case 2: Article Screening

Example 2: Screening and classification of grey literature using GenAI

a) Use Context: Grey literature review on the intersection between UX Design and Return on Investment (ROI).

b) Research Objective: To identify and select relevant documents from grey literature sources (e.g., blogs, white papers, market reports) that explicitly address both User Experience Design (UX Design) and Return on Investment (ROI).

c) Input Source: Full-text documents collected from public websites, online platforms, and organizational repositories containing professional and market-based publications.

d) Tools/Platforms: ChatGPT-4 (OpenAI), accessed via the ChatGPT web interface.

e) Prompt Role: You are a UX professional with strong knowledge of business impact metrics. Your role is to evaluate whether the provided text discusses both UX Design and ROI in a clear and relevant manner.

f) Task to perform: Read the document and classify it as relevant or not based on the presence of both UX Design and ROI. Justify your response with direct quotations.

g) Prompt: *Chat, read the full content of this grey literature document and indicate whether it addresses both UX Design and ROI. Reply with "YES"or "NO", and provide direct excerpts from the text that justify your classification. Only answer "YES"if both terms are discussed explicitly in the context of digital products or services.*

h) Output Explanation: ChatGPT returns a binary response (YES/NO) along with quoted excerpts that support the classification. Researchers validate these outputs through manual sampling checks.

i) Output Example:
YES.
Excerpt 1: "Investing in UX early in the development cycle leads to measurable ROI through reduced support costs and improved customer retention."
Excerpt 2: "This white paper demonstrates how UX improvements increased conversion rates, contributing to revenue growth."

4.3. Case 3 - Extraction with Predefined Codes

In this case study, conducted as part of one doctoral research, the **research goal** was to leverage ChatGPT for extracting relevant excerpts from articles, following a set of predefined codes. To assist in the development of this task, the **GenAI tools** utilized were the ChatGPT interface and via OpenAI API [5]. The **outputs** resulting from the interaction with ChatGPT, in the closed coding context, were mostly relevant excerpts consistent with the previously defined codes, contributing to the organization and analysis of the data. The **researcher experience** in using the ChatGPT for extracting excerpts based on predefined

[5]https://openai.com/api/

codes presents a balance between structured efficiency and the initial investment required. While the use of codes leads to organized and predictable output, it simplifies evaluation and can potentially save time in large-scale analysis. However, achieving this relies heavily on the researcher's effort in designing clear prompts and accurate codes. Poorly defined codes can lead to irrelevant extractions, necessitating careful upfront work to unlock the benefits of targeted information retrieval. The developed prompt is shown below.

Prompt case 3: Extraction with Predefined Codes

a) Use Context: Thematic analysis of academic articles in the field of Human-Computer Interaction and Smart Cities.

b) Research Objective: To extract and organize relevant excerpts from a set of selected articles based on predefined analytical codes.

c) Input Source: Full-text peer-reviewed articles previously selected through systematic review and provided in structured digital format.

d) Tools/Platforms: ChatGPT-4 (OpenAI), accessed via both the ChatGPT web interface and the OpenAI API.

e) Prompt Role: You are a senior researcher conducting a thematic analysis. Your role is to identify and extract verbatim excerpts from the articles that match specific, predefined codes.

f) Task to perform: Read the article and extract all passages that align with the provided code definition and examples.

g) Prompt:
Code: B. Context of use.
Description: Components of the context of use of the environment considered in the design process (e.g., technologies, sensors, mobile devices).
Example: Excerpts that describe challenges in urban environments or interactions between citizens and technologies.
Example excerpts:

- "The urbanization of cities and the need to provide quality services to citizens have become major challenges for governments."
- "ClouT focuses on leveraging Cloud computing to bridge IoT and people through services."
- "As such, the U-city project mainly considered urban infrastructure using information platforms."
- "NOMAD is a platform for policymakers to monitor and analyze citizens' discussions on social media."
- "The framework promotes the development of a city commons using accessible resources like data and technology."

h) Output Explanation: ChatGPT returns a list of excerpts from the article that match the code definition.

i) Output Example:
B. Context of Use
Excerpts:
- "The monitoring and control of all this activity is essential for its proper functioning. In this context, Information and Communication Technology (ICT) plays a key role in the digitisation, monitoring, and managing of these different verticals."
- "Vitoria-Gasteiz did not have an urban platform. The city had several monitoring systems for urban elements and services, which were used and maintained by the concessionaire of each urban service."

4.4. Case 4 - Inductive Open Coding

Conducted within a doctoral project, this case aimed to classify the studies identified in the systematic review by applying open coding to organize and label the evidence. The **research goal** of this case was to inductively identify new codes and perform extractions. The interface of **GenAI tools** and their APIs assisted the researcher on this task. The **outputs** resulting from the interaction with ChatGPT showed that the attempt to perform open coding inductively was not successful, as the model had difficulty identifying new codes with consistency and analytical depth. The **researcher experience** revealed important insights. Despite the initial appeal of the ChatGPT's flexibility for inductive code identification, open coding posed challenges for the model due to its inherent ambiguity. The outputs resulting from the interaction with ChatGPT showed that the attempt to perform open coding inductively was not successful. The model had

difficulty identifying new codes with consistency and analytical depth. This limitation led to a significant investment of the researcher's time and effort in prompt engineering, with minimal return. Ultimately, this highlighted the limitations of ChatGPT in handling unstructured, high-level analytical tasks without predefined guidance. The outputs were often unreliable and unsatisfactory. The developed prompt is shown below.

Prompt case 4: Inductive Open Coding

a) Use Context: Systematic Literature Review in HCI and Smart Cities.

b) Research Objective: To explore how technologies are integrated into co-design processes, with a focus on human participation and technical aspects.

c) Input Source(s): 53 full-text PDF articles from a systematic review.

d) Tools/Platforms: OpenAI API.

e) Prompt Role: Senior researcher in HCI conducting a literature review.

f) Task to perform: Inductively generate labels for relevant excerpts without predefined categories according to [Hitch 2024].

g) Prompt: You are a senior researcher in Human-Computer Interaction and want to review papers. You are using the open coding technique to identify labels that introduce meaning to excerpts of the text. Labels must identify the strategies (techniques, apparatus) used to integrate components, such as sensors, actuators, and mobile devices, into the design process of smart city solutions.
Analyze the article paragraph by paragraph and assign labels that add meaning to the excerpts of the paragraph. Use many labels as necessary to add a high level of abstraction to the meaning. Indicate which excerpts the labels correspond to. The excerpts must explain which components and elements of the context of use are involved, such as users, tasks, technologies, physical environment, data, and information.
Do not analyze sections with titles such as "Related Works,Literature Review,Theoretical Background,State of the Art, Background,Review of the Literature,Research Background,Previous Studies,Existing Literature,Knowledge Base,Foundation of the Study,Research Context,"or "Analysis of Related Research."
Instructions:
- Respond only with raw text excerpts from the article that illustrate the identified labels.
- Do not provide summaries, interpretations, or explanations.
- Extract the text exactly as it appears in the article, preserving the original structure, grammar, and phrasing.
- You can find more than one excerpt per label.
- If no labels are identified in certain parts of the article, omit those sections from the response.

h) Output Explanation: The output of the prompt consisted of a list of codes generated by the model, with each code followed by the corresponding excerpts extracted from the articles.

i) Output Example:
1. Citizen Engagement and Co-Creation
- "The urban strategy relies on a digital transformation, but considering the citizens as an important player (step "engage" of SZCC). Hence, engagement should also be applied on the development of the CIOP. This section presents the co-creation methodology."
- "An important aspect to consider when developing smart city solutions is to involve citizens in their design, testing, and validation."
2. Sustainability and Energy Efficiency
- "The main actuation in this way was the retrofitting of buildings to increment the energy efficiency."
3. Interoperability and Data Integration
- "CIOP is at the core for digital transformation of the lighthouse cities. The ICTs are thus considered as cross-cutting enabling technologies used for monitoring and evaluation of the success of measures."

4.5. Case 5 - Word Count Reduction

The **research goal** of this case was to use ChatGPT to reduce the word count of specific paragraphs while preserving their meaning and the arguments presented in the text. The mentioned texts were part of papers written in the context of a doctoral project on UX Research, Long-Term UX, and software startups. To assist in the development of this task, the **GenAI tool** utilized was ChatGPT with Write for Me. The **outputs** resulting from the interaction with ChatGPT included the revised version of the paragraph with the word count reduction applied, along with a report explaining how each specific

sentence was modified or rewritten. The **researcher experience** with ChatGPT for word count reduction in scientific articles reveals a helpful tool contingent on precise communication. While ChatGPT offers the potential to streamline the process of condensing text, especially in sections with word limits, the researcher quickly learns the necessity of providing specific instructions regarding the target word count and the preservation of meaning and citations. Before developing the prompt, the researcher should determine how many words are in the paragraph they want to shorten, which they can do using text editors. Then, to develop the prompt, the researcher should specify to ChatGPT the number of words in the paragraph and the number of words they want to remove. When constructing the prompt, the researcher should also clarify that reducing the word count must not affect the meaning of the arguments. An example of a prompt using this strategy is shown below.

Prompt case 5: Word Count Reduction

a) Use Context: Development of a scientific paper on UX Research practices and Long-Term UX cycle.

b) Research Objective: To investigate what UX Research practices software startups apply and how this work is connected with the Long-Term UX cycle.

c) Input Source: A paragraph originally written by the author, in addition to specifying the number of words in the paragraph, and the number of words they want to remove.

d) Tools/Platforms: ChatGPT from Write for Me.

e) Prompt Role: You are a UX researcher with strong academic experience in Human-Computer Interaction (HCI), qualitative research methods, and UX Research practices with a longitudinal perspective (Long-Term UX). Your task is to reduce the word count of paragraphs while preserving their meaning and the arguments presented in the text.

f) Task to perform: Reduce a specific number of words in each paragraph without changing the meaning of the arguments or references.

g) Prompt: *Chat, the paragraph below contains 198 words. I would like you to reduce it by 70 words, resulting in a total of 128 words. Please maintain the meaning of the arguments presented, as well as the integrity of any citations from other authors. Apply these instructions to the following paragraph:* [paragraph originally written by the author].

h) Output Explanation: ChatGPT provides the revised version of the paragraph with the word count reduction applied, along with a report explaining how each specific sentence was modified or rewritten.

i) Output Example: [new paragraph with 128 words and specifics change]
Changes made:
1. “we took care to maintain an approximate number of questions not to unbalance the amount of data obtained in each interview"→ “while maintaining a consistent number of questions to balance data collection"
- I rephrased it for brevity.
2. “not to take or stress out the participants"→ “to avoid participant fatigue"
- I simplified it for clarity.
‘3. ‘to limit any influence of individual opinion"→ “to reduce individual bias"
- I summarized keeping the meaning.
4. “from which the first researcher could review the data as often as they wished"→ “for thorough analysis"
- I summarize by eliminating redundancies.

4.6. Case 6 - Broad Information Extraction

The **research goal** of this case was to utilize GenAI for the analysis of scientific articles (including PDFs) to extract pertinent information for literature review and to pinpoint key findings, arguments, and topics relevant to developing the discussion of a study. The papers were selected in the context of a doctoral project about UX Research, Long-Term UX, and software startups. To assist in the development of this task, we used the **GenAI tool**: ChatGPT from Write for Me. The **outputs** expected from the interaction with ChatGPT include a topic-centered summary of the article and relevant excerpts

that address or support the searched topic, ultimately contributing to a streamlined literature review process. The **researcher experience** utilizing ChatGPT for analyzing and extracting information from scientific articles offers a significant boost to literature review efficiency and argument organization, especially with large volumes of text. However, the effectiveness of this process is heavily dependent on the clarity and precision of the prompts provided. Furthermore, the tool's limitations in handling multiple topics across numerous articles simultaneously can necessitate a more iterative and potentially time-consuming approach. This requires the researcher to carefully design their queries to maximize the benefits of AI-assisted information retrieval. To develop the prompt, the researcher instructs ChatGPT to search for a specific topic in a scientific article. The researcher must also clearly state what they expect the ChatGPT to extract from the article about the chosen topic. This search may involve a specific excerpt from the article that addresses the topic, or a summarized interpretation by the ChatGPT explaining how the article relates to the topic. An example of a prompt using this strategy, along with the response generated by ChatGPT after executing this prompt about the publication of [Hokkanen e Väänänen-Vainio-Mattila 2015], is shown below.

Prompt case 6: Broad Information Extraction

a) Use Context: Development of a scientific paper on UX Research practices and Long-Term UX cycle.

b) Research Objective: To investigate what UX Research practices software startups apply and how this work is connected with the Long-Term UX cycle.

c) Input Source: full-text PDF about a paper that the researcher wants to extract information.

d) Tools/Platforms: ChatGPT from Write for Me.

e) Prompt Role: You are a UX researcher with strong academic experience in Human-Computer Interaction (HCI), qualitative research methods, and UX Research practices with a longitudinal perspective (Long-Term UX). Your task is to analyze scientific papers to extract information such as key points, arguments, or findings about specific topics.

f) Task to perform: The task involves identifying key points, arguments, or findings in a scientific paper about one or more topics of interest to the researcher.

g) Prompt: *Chat, I want to know if the paper in the PDF presents results or discussions about the 'need for startups to set UX goals or objectives for software development'. If it does, summarize how this appears in the paper, or provide an excerpt from the paper that addresses the topic.*

h) Output Explanation: The response is presented with a summary about how the topic described in the prompt appears in the paper and, if possible, a snippet from the paper that proves the presence of the topic.

i) Output Example: In the paper "UX Work in Startups: Current Practices and Future Needs" by Hokkanen and Väänänen, it was possible to identify discussions and results about the need for startups to define UX objectives or goals in software development.
Summary: The paper mentions that, although some startups recognize the importance of defining clear goals for user experience (UX), this practice is not always systematized. Some startups reported having implicit goals, but few formally documented these intentions. One example is using user journey maps to plan specific emotions the product should evoke, demonstrating the attempt to align the design with clear emotional and functional goals. On the other hand, several startups adopted a more reactive approach, prioritizing development and collecting feedback later to adjust the user experience, rather than setting specific goals from the beginning.
Relevant quote: About startups and defining clear UX goals
Quote: "*Startup SU4 was the only startup that had clear UX goals. They had used the information gained from interviews to create user journey maps. They had defined emotions that the user should experience with the product, and the design was created to meet those goals.*"

4.7. Case 7 - Agreement Level in Screening and Selecting Studies

The **research goal** of this case was to leverage ChatGPT as a second reviewer in a literature review in a doctoral project on HCI to ensure the quality and rigor of

the article selection process by independently applying and verifying the predefined inclusion criteria. To assist in the development of this task, the **GenAI tool** utilized was the ChatGPT interface. The **outputs** resulting from the interaction with ChatGPT demonstrated its capacity to understand the research topic, providing coherent responses and relevant examples; the model efficiently selected articles based on title, keywords, and abstract, and to ensure the quality and rigor of this selection process, ChatGPT rated all articles, with this classification being compared with the author's evaluation. In order to verify the level of agreement of researchers for the selection of articles, we used the Kappa coefficient [Fleiss et al. 2003]. The level of agreement obtained between the researcher and ChatGPT was 0.62, which is classified as substantial agreement (0.61 – 0.80) [Fleiss et al. 2003]. This result supports the reliability of ChatGPT's application of the predefined inclusion criteria. The **researcher experience** using ChatGPT as a second reviewer for article selection is primarily characterized by its remarkable speed and demonstrated accuracy, significantly accelerating the screening process. However, the researcher encountered occasional instances where the model did not provide an immediate response, requiring a follow-up prompt. While the overall experience is positive due to the speed and reliability, the need for occasional re-prompting introduces minor interruptions to the workflow. The goal of the Prompt was to provide ChatGPT with context about the topics that will be addressed next. Additionally, it allows the researcher to verify whether ChatGPT's responses meet expectations and are aligned with the research objectives. An example using this strategy is shown below.

Prompt case 7: Agreement Level in Screening and Selecting Studies

a) Use Context: ChatGPT as a second reviewer in a literature review on Human-Computer Interaction.

b) Research Objective: Research integration approaches between UCD and Agile Development.

c) Input Source: Title, Abstract and Keywords of selected articles on the research area.

d) Tools/Platforms: ChatGPT.

e) Prompt Role: You are a researcher in Human-Computer Interaction who is conducting a literature review.

f) Task to perform: Evaluate whether the paper (based on the title, abstract and keywords) is related to the research area of the literature review. Based on this evaluation, return a score from 1 to 7 (where 1 = strongly disagree and 7 = strongly agree).

g) Prompt: Assume you are a Human-Computer Interaction researcher conducting a systematic literature review (SLR). Consider the title, abstract, and keywords of a primary study.
- Title: [the title of the primary study]
- Abstract: [the abstract of the primary study]
- Keywords: [the keywords of the primary study]
- Using a 1-7 Likert scale (1 - Strongly disagree, 2 - Disagree, 3 - Somewhat disagree, 4 - Neither agree nor disagree, 5 - Somewhat agree, 6 - Agree, and 7 - Strongly agree) rate agreement with the following statement (only number): Papers that describe integration approaches between UCD and Agile Development. ChatGPT, could you help me with the question above?

h) Output Explanation: The output of the prompt in ChatGPT was a brief justification and the evaluation score, in the case of the example, rating 7, which means (strongly agree).

i) Output Example: Based on the title, abstract, and keywords of the primary study provided, I would rate my agreement with the statement "Papers that describe integration approaches between UCD and Agile Development"as 7 (Strongly agree).

4.8. Case 8 - Idea and Subtopic Generation

The **research goal** of this case was to leverage GenAI for the generation of relevant subtopics and initial ideas within a specific theme (e.g., UX in industry). The sentences were written in the context of a doctoral project about UX Data, and Information

Visualization. To assist in the development of this task, the **GenAI tool** utilized was the ChatGPT. The **outputs** from the interaction with ChatGPT included a list of relevant subtopics, rationales for each subtopic, and expanded discussion on selected subtopics. The **researcher experience** using ChatGPT for generating subtopics and initial ideas offers a valuable starting point for organizing content and ensuring key elements are considered. The tool encourages systematic thinking and can improve writing coherence. However, the quality of the generated output is heavily reliant on the clarity of the initial prompt. The researcher may find that the ChatGPT produces generic or overly broad suggestions, which require further refinement and domain-specific knowledge to become truly useful for their research. An example of a prompt using this strategy is shown below.

Prompt case 8: Idea and Subtopic Generation

a) Use Context: Scientific article about UX data.

b) Research Objective: To investigate how UX data is used in practice, focusing on its characteristics, and motivations for exploration.

c) Input Source: The central theme of the paragraph that will be written.

d) Tools/Platforms: ChatGPT.

e) Prompt Role: You are a UX researcher with strong academic experience in Human-Computer Interaction (HCI), qualitative research methods, and UX data analysis. Your task is to evaluate research papers, extract structured metadata, and help rephrase content for academic clarity.

f) Task to perform: Generate a list of relevant subtopics, provide rationales for each, and expand on selected ideas.

g) Prompt: *Hello ChatGPT, could you help me identify key topics to cover in a paragraph about using UX in industry?*

h) Output Explanation: The response is structured with a brief introductory framing that contextualizes the intent of the list. Then a sequential list of categorized key points. Each item maintains a consistent template (i.e., a bolded concept followed by a brief explanatory phrase) enhancing readability and clarity.

i) Output Example: Absolutely! Here's a list of key topics you might consider when structuring a paragraph about UX in industry—designed to help guide your own writing: Role of UX in aligning user needs and business goals: Emphasize UX as a strategic asset, not just a design function. Integration of UX into product development cycles: Consider how UX fits into agile, lean, or iterative workflows. Methods used in industry: Mention common qualitative (e.g., interviews, usability testing) and quantitative (e.g., analytics, A/B testing) approaches. UX data practices: Highlight how UX data is collected, interpreted, and used to support decisions. Cross-functional collaboration: Note the interaction between UX researchers, designers, developers, and product managers. Constraints and challenges: Include industry limitations such as time pressure, limited access to users, or stakeholder alignment. Value and impact of UX: Touch on outcomes like improved user satisfaction, reduced support costs, or increased product adoption. Long-term UX practice: Reflect on sustaining UX as an ongoing process rather than a one-time intervention.

4.9. Case 9 - Data Visualization Suggestions

The **research goal** of this case was to enhance the clarity and impact of research findings by generating suggestions for appropriate data visualizations based on text content. The visualizations were developed in the context of a doctoral project about UX Data, and Information Visualization. To assist in the development of this task, the **GenAI tool** utilized was the ChatGPT. The **outputs** resulting from the interaction with ChatGPT included suggested types of data visualizations (e.g., bar charts, line graphs, pie charts), descriptions of how each visualization relates to the data/content, and enhanced presentation of complex information. The **researcher experience** using ChatGPT for data visualization suggestions reveals a helpful tool for structuring complex data and potentially improving the clarity of findings communication, while also saving time in deciding on appropriate chart formats. However, the quality of these suggestions is contingent on the level of detail provided in the input. The ChatGPT's output is limited

to textual descriptions of visualizations, requiring the researcher to use separate tools to create the actual graphics. Therefore, the researcher benefits from guidance but still needs to apply their contextual knowledge and utilize visualization software for implementation. An example of a prompt using this strategy is shown below.

Prompt case 9: Data Visualization Suggestions

a) Use Context: Scientific article about UX data.

b) Research Objective: To investigate how UX data is used in practice, focusing on its characteristics, and motivations for exploration.

c) Input Source(s): A paragraph containing qualitative data and conclusions, originally written for the results section of the article.

d) Tools/Platforms: ChatGPT.

e) Prompt Role: You are a UX researcher and data visualization expert with strong academic experience in Human-Computer Interaction (HCI), qualitative research methods, and UX data analysis. Your task is to support the design of clear, meaningful, and publication-ready charts and diagrams for academic papers.

f) Task to perform: The task involves identifying and match each visualization format to specific analytical goals. The response should consider the structure of the data (categorical, multi-dimensional) and communicate visual strategies that enhance interpretability and insight extraction.

g) Prompt: Hi ChatGPT, can you suggest some charts based on this content: *"Considering the context of the company where they worked, participants should list which professionals could obtain meaningful insights using the study's visualizations or whether the visualizations would not be helpful. All the visualizations received at least one vote as not applicable. In this question, participants should choose only one option. The data shows that in three of the visualizations, the participants were divided into two roles as being those most interested in using the visualization as a helpful tool; this division only did not occur in the "Interaction with specific user interface" visualization (i.e., which had the majority of votes directed to designer) and in the "Most searched terms"visualization (i.e., which was understood as useful to stakeholders). Designers and stakeholders were chosen as preferred for the "Last pages visited" and "Time in the system" visualization, while developers and stakeholders were appointed for the "Most searched terms"visualization"*. Please respond in a concise textual way without bullet points.

h) Output Explanation: The response adopts a structured, multi-option format to support visualization selection. It begins with a default recommendation (i.e., grouped bar chart), while subsequent alternatives are introduced through conditional language (e.g., "alternatively", "if you want to emphasize") to help researchers align each visualization with specific analytical objectives. Each suggestion follows a consistent template: the visualization type is named, followed by a concise description of its layout and interpretive function.

i) Output Example: You can represent this data effectively using a grouped bar chart, where each visualization appears on the x-axis and the number of participants per professional role (designer, developer, stakeholder, or "not applicable") is shown as grouped bars. This format clearly displays role preferences and highlights divisions or consensus. Alternatively, a heatmap matrix with visualizations as rows and roles as columns can offer a compact view of selection intensity, with a distinct marker for "not applicable" cases. If you want to emphasize the flow of preferences, a Sankey diagram mapping visualizations to roles can illustrate shared interest, especially where two roles dominate a single chart. For highlighting polarizing responses, a diverging stacked bar chart can separate useful from "not applicable" votes visually.

5. Lessons Learned

By analyzing the case reports, we identified six emerging themes as the basis of our findings. The identified themes include *refining the content, shortening text, identifying key points for insights, classifying for screening, summarizing studies with key-point extraction,* and *assisting with data visualization*. Considering the objectives of each case, we discuss the strengths and weaknesses of the approaches adopted by the researchers and present them in the format of lessons learned.

Lesson 1 - Refining the content. This lesson is related to the Text Review case (Case 1) that revealed the practical utility of ChatGPT for a specific and time-consuming aspect of academic research writing refinement, particularly benefiting non-native English

speakers. However, while it significantly reduces the time spent on reading and correcting, we noted that achieving a deeper understanding of the changes requires explicit prompting for justifications. Furthermore, the recommendation to revise paragraph by paragraph, to mitigate hallucination risks, introduces a more granular approach that can partially offset the initial time savings, necessitating a strategic and mindful engagement with the AI to maximize its benefits and ensure reliable results. In summary, this lesson highlights the need for explicit prompting to gain deeper insights into the ChatGPT's corrections and the importance of a cautious, segmented approach to mitigate potential inaccuracies.

Lesson 2 - Shortening the text. This lesson addresses the use of a ChatGPT for reducing the word count (Case 5) when the researcher needs to respect a maximum number of words to submit a scientific article. Sections such as related work and discussions in a manuscript can become leaner and more objective with the support of ChatGPT. However, the experience reported in this activity has shown that ChatGPT may not provide satisfactory answers when the researcher fails to specify the exact number of words in the current paragraph and how many words they intend to reduce. It is also necessary to make it clear to ChatGPT that the reduction of words should not affect the meaning of the arguments developed by the researcher regarding the citation of other authors, to limit the creativity of the ChatGPT in the answer. The recommendation to process text paragraph by paragraph also introduces a more methodical approach, balancing the time-saving aspect with the need for careful, segmented interaction to ensure accurate and reliable results.

Lesson 3 - Identifying key points for insights. The lesson on idea and subtopic generation (Case 8) revealed that researchers can experience a boost in systematic thinking and writing organization when using ChatGPT to identify insightful information on their research concerns. The GenAI tools help improve coherence and flow in writing by providing an initial structure that can guide content development. In addition, using ChatGPT provides a safety net, helping to ensure that important content elements are not overlooked in the planning and ideation phase. This ability to structure content upfront can be particularly useful for researchers who are starting a new project or exploring a complex topic. However, experience also shows that the quality of the output is intrinsically linked to the clarity of the prompt provided. A vague or poorly defined prompt can lead to generic results or overly broad subtopics that will require further refinement by the researcher to become truly useful and specific to their research. This means that the researcher needs to invest time and effort in crafting precise prompts to drive ChatGPT effectively.

Lesson 4 - Classifying for screening. A cross-case lesson emerging from four cases (Cases 2, 3, 4 e 7) is that ChatgPT performs better in literature-review tasks when its behavior is anchored to explicit, highly structured instructions - such as predefined inclusion/exclusion criteria or closed code lists — yet falters when asked to operate inductively without that scaffold. In the screening cases, the model acted as a second reviewer, quickly classifying papers, justifying each YES/NO decision with verbatim evidence, and achieving substantial agreement with humans. In closed coding it likewise pulled highly relevant excerpts, accelerating synthesis. These strengths translate into time-savings and heightened transparency that can increase the overall rigor of a review. Conversely, the open-coding attempt exposed the model's limitations:

ambiguity about what constitutes a "new" code led to inconsistent, shallow results, while even the structured tasks revealed risks of silent omissions, occasional non-responses, and erroneous exclusions whenever the prompt or source text was unclear. Mitigation therefore, hinges on segmenting large documents, iteratively refining prompts, and maintaining active human oversight to catch ambiguities—demonstrating that clear structure is the keystone for reliable ChatGPT support in systematic review workflows. A further practical insight, observed in Case 3, is that lowering the model's temperature (i.e., reducing sampling randomness) markedly reduced hallucinations and produced extractions that matched the source text verbatim. Thus, conservative temperature settings—used alongside structured prompting—constitute an effective, low-effort lever for improving fidelity and trustworthiness in ChatGPT-assisted evidence extraction.

Lesson 5 - Summarizing studies with key-point extraction. This lesson derives from the Broad Information Extraction exercise documented in Case 6, where GenAI (ChatGPT, via Write for Me) was employed to analyse scientific articles—including PDFs—and generate topic-centred summaries accompanied by verbatim excerpts that substantiate each key point. The strategy markedly accelerated the literature-review workflow by foregrounding pertinent findings, arguments, and themes, thereby enabling researchers to organise evidence and construct well-rounded discussions even when dealing with extensive, multi-topic corpora. Nevertheless, its success proved highly contingent on precise, well-scaffolded prompting: vague or ambiguous requests often produced incomplete or confusing extractions. Moreover, when numerous topics or large sets of papers were processed simultaneously, the system required a laborious, iterative prompting cycle, underscoring the necessity of critical human oversight to validate relevance and accuracy and to refine queries for optimal, context-specific retrieval.

Lesson 6 - Assisting with data visualization. This lesson is grounded in the Data Visualization Suggestions experience (Case 9), which demonstrated the practical value of ChatGPT in expediting an otherwise demanding stage of disseminating empirical results—namely, selecting appropriate graphic formats for complex datasets. The model's rapid, text-based recommendations (e.g., advocating grouped bar charts for categorical contrasts or line graphs for temporal trends) substantially shorten the deliberation period for researchers who lack deep expertise in visual analytics. However, while these suggestions accelerate initial decision-making, we observed that their contextual accuracy is highly contingent upon the granularity of the prompt: insufficiently detailed inputs often yield generic or misaligned proposals. Moreover, the necessity to translate ChatGPT's purely narrative guidance into concrete figures within external visualisation software re-introduces a manual step that partially offsets the time savings realised during the advisory phase. Consequently, researchers must engage the ChatGPT with richly specified descriptions of variables, audience, and intended insights, and then employ a complementary toolchain to instantiate the recommended visuals. In summary, this lesson underscores both the efficiency gains attainable through ChatGPT-assisted chart selection and the critical importance of precise prompting and downstream software integration to ensure contextually valid and practically usable visualisations.

6. Final Remarks

This study aimed to investigate how Generative Artificial Intelligence (GenAI) tools – especially large language models like ChatGPT – are being strategically employed in

Human-Computer Interaction (HCI) research. Based on a qualitative analysis of nine distinct cases, using both closed and open coding, six lessons learned were identified, illustrating the potential of these technologies to support tasks such as text refinement, article screening, information extraction, and data visualization suggestions. The findings show that, when consciously integrated into research workflows, these tools can increase productivity and enhance the quality of analytical outputs in the HCI field.

Despite the advantages observed, the results also reveal important limitations, particularly in tasks that require open-ended interpretation and critical judgment. ChatGPT tools proved to be more effective in structured and well-defined activities, while their performance was less consistent in exploratory and interpretive tasks. This reinforces the irreplaceable role of the human researcher, whose expertise is essential both to guide the prompts provided to the tool and to validate and critically interpret its outputs. In short, ChatGPT should be understood as a strategic assistant capable of accelerating operational stages but not replacing human reasoning and analysis. Furthermore, it is worth noting that although the activities with GenAI were carried out in the context of HCI research, we believe that such activities can be extended and bring benefits to other areas of research.

We conclude that the reflective integration of ChatGPT into HCI research processes represents an opportunity to combine efficiency and innovation, as long as it is mediated by critical and ethical engagement from researchers. This study contributes by offering a practical and grounded overview of the possible uses of ChatGPT across different stages of the research cycle, serving as a reference for future initiatives in the field. This study was conducted within the context of our research group, which may influence its results and limit their generalizability to other contexts. Future research may explore other contexts and more sophisticated models of human-AI collaboration, focusing on transparency, reliability, and the expansion of research practices in diverse HCI contexts.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Finance Code 001), Consellho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Brazil) (grant 309497/2022-1), and by São Paulo Research Foundation (FAPESP) (grant 2020/11441-1). We also extend our thanks to the CONECTA2030 project (FUNDEP - process number 29271.02.01/2022.04-00), and IFSULDEMINAS for its contribution to the development of this research. The authors declare that they wrote this article and used Artificial Intelligence tools to support translation and grammar checking.

CRediT authorship contribution statement

Tatiana Alencar: Conceptualization, Writing - original draft, Writing - review & editing. **Leonardo Ciscon**: Conceptualization, Writing - original draft, Writing - review & editing. **Gessé Pereira**: Conceptualization, Writing - original draft, Writing - review & editing. **Maylon Macedo**: Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Suéllen Martinelli**: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Joelma Choma**: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Luciana Zaina**: Validation, Writing - review & editing, Supervision.

Referências

Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A. E., e Zayed, T. (2023). Harnessing the power of chatgpt for automating systematic review process: methodology, case study, limitations, and future directions. *Systems*, 11(7):351.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Burger, B., Kanbach, D. K., Kraus, S., Breier, M., e Corvello, V. (2023). On the use of ai-based tools like chatgpt to support management research. *European Journal of Innovation Management*, 26(7):233–241.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., et al. (2024). A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, 15(3):1–45.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications.

Chen, C., Lee, S., Jang, E., e Sundar, S. S. (2024). Is your prompt detailed enough? exploring the effects of prompt coaching on users' perceptions, engagement, and trust in text-to-image generative ai tools. In *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems*, pages 1–12.

Dong, M. M., Stratopoulos, T. C., e Wang, V. X. (2024). A scoping review of chatgpt research in accounting and finance. *International Journal of Accounting Information Systems*, 55:100715.

Fabiano, N., Gupta, A., Bhambra, N., Luu, B., Wong, S., Maaz, M., Fiedorowicz, J. G., Smith, A. L., e Solmi, M. (2024). How to optimize the systematic review process using ai tools. *JCPP advances*, 4(2):e12234.

Felizardo, K. R., Lima, M. S., Deizepe, A., Conte, T. U., e Steinmacher, I. (2024). Chatgpt application in systematic literature reviews in software engineering: an evaluation of its accuracy to support the selection activity. In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pages 25–36.

Fleiss, J. L., Levin, B., e Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc, 3rd edn. edition.

Freire, A. P., Cardoso, P. C. F., e Salgado, A. d. L. (2023). May we consult chatgpt in our human-computer interaction written exam? an experience report after a professor answered yes. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–11.

Gibbs, G. (2018). *Analyzing Qualitative Data*. SAGE Publications Ltd, 2 edition.

Gilardi, F., Alizadeh, M., e Kubli, M. (2023). Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Hanafi, A. M., Al-mansi, M. M., Al-Sharif, O. A., et al. (2025). Generative ai in academia: A comprehensive review of applications and implications for the research process.

International Journal of Engineering and Applied Sciences-October 6 University, 2(1):91–110.

Hitch, D. (2024). Artificial intelligence augmented qualitative analysis: the way of the future? *Qualitative Health Research*, 34(7):595–606.

Hokkanen, L. e Väänänen-Vainio-Mattila, K. (2015). Ux work in startups: Current practices and future needs. In Lassenius, C., Dingsøyr, T., e Paasivaara, M., editors, *Agile Processes in Software Engineering and Extreme Programming*, pages 81–92, Cham. Springer International Publishing.

Hosseini, M. e Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? considerations and recommendations for use of chatgpt and other large language models in scholarly peer review. *Research integrity and peer review*, 8(1):4.

Khalifa, M. e Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, page 100145.

Lazar, J., Feng, J. H., e Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, 2nd edn. edition.

Lefebvre, C., Manheimer, E., e Glanville, J. (2008). Searching for studies. *Cochrane handbook for systematic reviews of interventions: Cochrane book series*, pages 95–150.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., e Moher, D. (2009). The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj*, 339.

Liu, V. e Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–23.

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., e Wang, Z. (2023). Chatgpt and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5):570–581.

Martinelli, S., Choma, J., e Zaina, L. (2023). Análise Qualitativa em IHC: da codificação à criação de visualizações. In *Minicursos do XXII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 24–45, Porto Alegre, RS, Brasil. SBC.

Perkins, M. e Roe, J. (2024). Generative ai tools in academic research: Applications and implications for qualitative and quantitative research methodologies. *arXiv preprint arXiv:2408.06872*.

Petersen, K. e Gerken, J. M. (2025). On the road to interactive llm-based systematic mapping studies. *Information and Software Technology*, 178:107611.

Rahman, M. M., Terano, H. J., Rahman, M. N., Salamzadeh, A., e Rahaman, M. S. (2023). Chatgpt and academic research: A review and recommendations based on practical examples. *Rahman, M., Terano, HJR, Rahman, N., Salamzadeh, A., Rahaman, S.(2023). ChatGPT and Academic Research: A Review and Recommendations Based*

on Practical Examples. Journal of Education, Management and Development Studies, 3(1):1–12.

Schmidt, D. C., Spencer-Smith, J., Fu, Q., e White, J. (2024). Towards a catalog of prompt patterns to enhance the discipline of prompt engineering. *ACM SIGAda Ada Letters*, 43(2):43–51.

Tenhundfeld, N. L. (2023). Two birds with one stone: Writing a paper entitled "chatgpt as a tool for studying human-ai interaction in the wild" with chatgpt. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 67, pages 2007–2012. SAGE Publications Sage CA: Los Angeles, CA.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., e Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Watkins, R. (2024). Guidance for researchers and peer-reviewers on the ethical use of large language models (llms) in scientific research workflows. *AI and Ethics*, 4(4):969–974.

Wen, J. e Wang, W. (2023). The future of chatgpt in academic research and publishing: A commentary for clinical and translational medicine. *Clinical and Translational Medicine*, 13(3):e1207.

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., e Oudeyer, P.-Y. (2023). Supporting qualitative analysis with large language models: Combining codebook with gpt-3 for deductive coding. In *Companion proceedings of the 28th international conference on intelligent user interfaces*, pages 75–78.

Zaina, L., Prates, R. O., Delabrida Silva, S. E., Choma, J., Valentim, N. M. C., Frigo, L. B., e Bicho, A. D. L. (2024). Grandihc-br 2025-2035-gc7: Interaction with emerging technologies: An ecosystem integrating humans technologies and contexts. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–21.