Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

PDF Download
3757534.pdf
27 December 2025
Total Citations: 0
Total Downloads: 223

Latest updates: https://dl.acm.org/doi/10.1145/3757534

Published: 16 October 2025

Citation in BibTeX format

RESEARCH-ARTICLE

'HistoChat': Leveraging AI-Driven Historical Personas for Personalized and Engaging Middle School History Education

YEON SOO KIM, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

HYUN-SEUNG MOON, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

SANG-SU LEE, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

TAKYEON LEE, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

Open Access Support provided by:

Korea Advanced Institute of Science and Technology

Proceedings of the ACM on Human-Computer Interaction, Volume 9, Issue 7 (October 2025)
https://doi.org/10.1145/3757534
EISSN: 2573-0142

“HistoChat”: Leveraging AI-Driven Historical Personas for Personalized and Engaging Middle School History Education

YEON SOO KIM*, KAIST, Republic of Korea
HYUN SEUNG MOON*, KAIST, Republic of Korea
SANGSU LEE, KAIST, Republic of Korea
TAK YEON LEE, KAIST, Republic of Korea

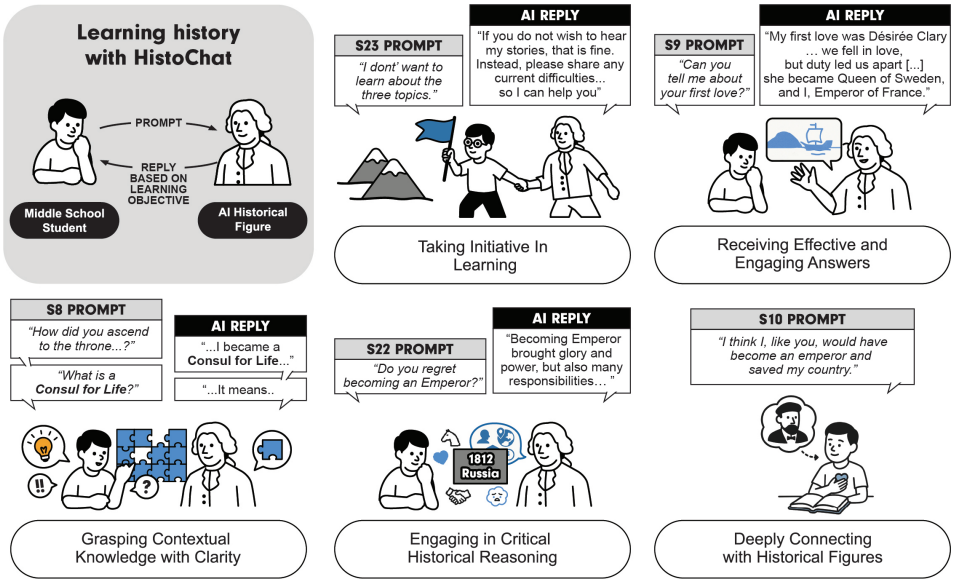

Fig. 1. Overview: Flowchart of HistoChat interaction(top-left) and the five benefits of using HistoChat.

Traditional history education often fails to cultivate historical empathy due to rigid curricula and limited opportunities for personalized, emotionally resonant engagement. We explore the potential of LLM-based historical personas to address these gaps by enabling students to engage in real-time, conversational interactions with simulated historical figures. A formative study with teachers and students surfaced key challenges and expectations around AI-mediated historical dialogue, informing the development of *Baseline* and *Experimental* HistoChat, AI persona systems featuring differing prompting strategies. A subsequent user study showed that these interactions fostered deeper inquiry, curiosity, and emotional engagement—while also revealing key limitations. From a CSCW perspective, this work expands the role of AI from task assistant to epistemic

*Both authors contributed equally to this research.

Authors' Contact Information: Yeon Soo Kim, Department of Industrial Design, KAIST, Daejeon, Republic of Korea, ykim18@kaist.ac.kr; Hyun Seung Moon, Department of Industrial Design, KAIST, Daejeon, Republic of Korea, mzes0401@kaist.ac.kr; Sangsu Lee, Department of Industrial Design, KAIST, Daejeon, Republic of Korea, sangsu.lee@kaist.ac.kr; Tak Yeon Lee, Department of Industrial Design, KAIST, Daejeon, Republic of Korea, takyeonlee@kaist.ac.kr.

This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2025 Copyright held by the owner/author(s).
ACM 2573-0142/2025/11-ARTCSCW353
https://doi.org/10.1145/3757534

partner, contributing to ongoing discourse on how dialogic systems can support meaning-making, empathy, and co-constructed learning in educational settings. Our findings yield valuable insights into the impact of tailored AI interactions on personalized and empathetic history education.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Applied computing → Interactive learning environments**.

Additional Key Words and Phrases: AI persona, LLM, AI in education

ACM Reference Format:
Yeon Soo Kim, Hyun Seung Moon, Sangsu Lee, and Tak Yeon Lee. 2025. "HistoChat": Leveraging AI-Driven Historical Personas for Personalized and Engaging Middle School History Education. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW353 (November 2025), 37 pages. https://doi.org/10.1145/3757534

1 Introduction

As Edward Hallett Carr famously stated, "history is a continuous process of interaction between the historian and his facts, an unending dialogue between the present and the past" [10]. This view frames history not as a static fact but as an evolving conversation shaped by present-day perspectives. In education, this calls for a shift from rote memorization to interpretive and relational engagement. A widely adopted framework that resonates with this perspective is "Historical Empathy" [56]—the ability to understand, emotionally connect with, and contextualize the lived experiences of historical figures [20]. Through perspective-taking, students gain insight into not only historical individuals but also the broader political, economic, and cultural forces shaping their lives [56]. This process fosters connections between the past and the present [23], encouraging students to reflect on their own values and decisions as historically situated [56]. In this way, practicing historical empathy provides a pathway for enacting Carr's vision of history as an interpretive, dialogic engagement with the past and deepens understanding of history's complexity.

Despite its promise, historical empathy remains difficult to cultivate in traditional classrooms. Textbook-driven instruction and standardized assessments reduce history to a static timeline, limiting opportunities for interpretation, emotional resonance, or human connection [23]. Teachers struggle to move beyond presenting facts to help students see historical figures as complex individuals, while uniform teaching methods fail to accommodate students' diverse perspectives, interests, or prior knowledge [71]. Without opportunities for reflective inquiry or personalized exploration, many students experience history as distant and disengaging—offering little space to build the emotional and cognitive bridges that historical empathy requires.

In response to these limitations, educators have explored strategies that help students see the past through more human and emotionally resonant lenses. One such method is classroom role-playing, where students adopt the perspectives of historical figures to explore their motivations, dilemmas, and decisions [56]. These activities are known to foster empathy and expand understanding, yet they require significant teacher facilitation and rely heavily on peer collaboration [53]. To extend perspective-taking beyond logistical constraints, recent advancements in Artificial Intelligence (AI) and Large Language Models (LLMs) offer new possibilities for creating similarly immersive and scalable learning experiences [64].

Central to this potential is the use of AI personas—dialogue agents that do not simply provide answers, but instead simulate the voices, values, and worldviews of historical figures. By engaging students in direct, conversational interaction, these personas can help bridge the distance between learners and the past, supporting both emotionally resonant and contextually grounded encounters with the past [76]. Unlike traditional AI tutors, which focus on instruction, AI personas are designed to inhabit specific identities [12, 77], maintaining consistent character traits and offering personalized and relevant responses [87]. AI personas can function as consistent

and immersive partners in historical learning that can also adapt to students' questions, interests, and pace—addressing long-standing challenges in supporting diverse learners. In a subject where storytelling, perspective-taking, and contextual understanding are essential [83], such interactive personas offer a compelling way to cultivate historical empathy. By transforming static content into emotionally resonant dialogues, they enable students to explore multiple perspectives and build meaningful connections with the past—making history more dynamic, relatable, and personally meaningful.

Despite their promise, most research on AI in education has focused on tutor-style systems [55, 97, 98], prioritizing factual delivery over emotional or interpretive engagement. One exception involves a social bot designed for empathetic history learning [70]; however, the bot acted more as a facilitator rather than embodying a historical character—limiting immersion and direct engagement with the historical perspectives. And yet, a key gap remains: How do students interact with AI personas designed to explicitly simulate historical figures? Addressing this question is especially important, because shifting from instructional support to character-driven dialogue introduces fundamentally different conditions for cultivating historical empathy. Such empathy depends not only on content but also on the interpretive quality of the exchange [77], pointing to the importance of understanding how students perceive and respond to these interactions. Still, most prior studies have concentrated on measurable outcomes rather than experiences—offering little insight into how students actually engage, reflect, and connect through AI-mediated historical dialogue.

To address these gaps, this paper examines how AI personas can foster historical empathy through a user-centered, empirical design study. Specifically, we examine how AI-generated historical characters are perceived and experienced with respect to engagement, empathy, and learning in a middle school classroom. Our research questions are:

(1) How do students perceive the AI-powered historical personas, and how are these perceptions related to their engagement?
(2) How are two AI-powered historical personas with differing prompting strategies experienced and compared, and what qualities and drawbacks are associated with each?
(3) What benefits and educational value are attributed to the use of AI-powered historical personas in history learning?

To address these research questions, we first conducted a formative study using ChatGPT-4o to examine how students and teachers interacted with AI-powered historical figures in classroom settings. This initial exploration focused on identifying restraints in traditional learning, observing initial reactions to historical personas, uncovering practical hurdles in using AI, and deriving design implications for integrating AI personas into history education. Building on these findings, we developed two versions of HistoChat—*Baseline*, which offered reactive responses, and *Experimental*, which incorporated proactive responses. We then carried out a within-subjects mixed-method study to explore how these two versions shaped students' experiences and evaluations, with particular attention to how each supported emotional connection, critical engagement, and personalized learning.

This research contributes to the CSCW community in two ways. First, we offer a deeper understanding of how AI personas can address long-standing barriers to historical empathy and engagement. By exploring through alternative prompt engineering strategies, we provide insights into the design of conversational systems that connect students meaningfully with the past. Second, we expand CSCW's scope by shifting attention from task-based human-AI collaboration to affective, meaning-centered interaction. By framing AI as a dialogic partner who embodies historical perspectives, we explore the co-construction of empathy through conversation—opening new directions for educational, emotionally resonant human-AI interaction.

2 Related Works

2.1 Generative AI for Education

Generative AI (GenAI) offers significant potential in education by creating diverse content such as images, text, and music with human-like creativity and variation [64]. In particular, GenAI enables personalized learning experiences by tailoring course content and strategies to learners' behavior patterns and progress [37]. Educational platforms powered by AI can offer customized feedback, recommendations, and timely interventions to support individual learning journeys [1], promote personalized environments, and improve knowledge retention [39]. Chatbots are increasingly being used to provide personalized help and support to students. They can answer questions, offer guidance, and serve as valuable resources throughout the learning process [45, 91]. These AI-driven tools enhance student engagement by providing real-time help, making learning more accessible and responsive to individual needs. On the teachers' side, GenAI enables real-time feedback and assessment, which allows teachers to monitor student progress and make data-driven instructional decisions. Early identification of struggling students allows for timely interventions and targeted support, helping learners overcome challenges and reach their full potential [2].

Despite its potential, the use of GenAI in education raises several concerns about its negative impact on assessment practices, scientific integrity, and the development of higher-order thinking skills of learners [24]. Questions also persist regarding the reliability and validity of AI-generated content, which can affect its acceptance among educators and students [90]. Ethical challenges arise from the varying regulations and codes of conduct in different educational contexts, which complicate the responsible use of GenAI [38, 47]. Furthermore, there is no solid empirical evidence that shows that GenAI promotes creativity, collaboration, or inspiration in meaningful ways [72]. Another key concern is the lack of research on the effectiveness of GenAI in building essential higher-order skills, such as argumentation, critical thinking, and reasoning, which are crucial educational objectives [3, 61, 65].

A broad consensus is that GenAI in education presents both opportunities and challenges. Its benefits in personalized learning and adaptive teaching are evident, yet the complexities of its effective integration demand further investigation. The learning sciences community faces the challenge of using GenAI to improve learning outcomes while addressing associated risks and limitations.

2.2 Educational Chatbots and Prompt Engineering

In recent years, the growing popularity of instant messaging has motivated educators to integrate chatbots and pedagogical agents into teaching and learning. Although early evaluations [16] found them to be limited conversation partners, the latest LLM-powered chatbots have shown significant improvements and potential in distance and online education [32]. Many chatbots available on platforms like Messenger focus on topics such as languages and economics, although most still rely on decision trees rather than advanced AI [82]. Recently, LLM-powered chatbots have been developed for the training of teachers [52, 58] and for the education of students [14, 84, 88]. These tools can simulate traditional classroom interactions while enhancing the user experience through more dynamic engagement [99].

In many cases, chatbots play the role of automated personal tutors, democratizing one-on-one tutoring [13, 40, 68, 81]. Such chatbots can engage students in conversational interactions that closely resemble human tutoring [28], helping them develop their skills while maintaining motivation and focus on their goals [63]. In addition, by analyzing student responses, chatbots can optimize learning outcomes, with cognitive tutoring models that further improve efficacy

by adapting to learners' cognitive levels and personalizing feedback and task difficulty to meet individual needs [44].

Despite their potential, chatbots in education raise several concerns. Issues of reliability and accuracy are significant, as chatbots may provide biased responses or inaccurate information [43, 78]. While ChatGPT can offer engaging and thought-provoking answers, it should not be considered a reliable source of information [80]. Biases in training data may lead to skewed perspectives, stereotypes, or discriminatory language, posing risks in educational settings [49]. Ethical concerns include data privacy, security, and the responsible use of AI [59, 60, 78, 86]. Students have reported difficulties in formulating effective prompts, indicating that sufficient background knowledge is required to benefit from chatbot outputs [54]. Academic integrity is another concern, as students may rely on LLM-generated content instead of producing original responses [54]. Additionally, chatbots often generate overly long, complex explanations [96] and may lack the specificity needed to address educational tasks effectively [46].

In response to the above concerns, existing materials cover a wide range of high-level guidelines and prompt engineering techniques [26, 62, 66, 74]. However, it remains unclear whether these materials are used in practice, as students can avoid effective prompting strategies that feel unnatural, despite knowing their benefits [54, 96].

2.3 AI Role-Playing in Educational Domains

Learning is inherently a social process that engages multiple stakeholders, including students, educators, and the broader community [89]. Thus, existing materials for the effective use of educational chatbots list a wide range of roles that can be assigned to chatbots [62]. In the field of GenAI, there are two lines of research on using persona: (1) role playing and (2) personalization [87]. LLM role-playing is actively employed in software development [36], game [69], and medical domains [94]. However, most existing research in the education domain has focused on providing personalized tasks, such as providing detailed step-by-step explanations [29], addressing complex technical topics [4], and summarizing lectures to enrich the learning experience [27] - with a few exceptions as below.

Lieb et al. [54] implemented two distinct modes for their educational chatbot by appending an unseen "system message" to each dialogue, defining the chatbot's role and behavior. In Tutor mode, the chatbot adopts a tone appropriate for the student's level, uses explanations, examples, and analogies, avoids giving complete solutions upfront, and encourages learning through questions. In Feedback mode, the chatbot provides contextual feedback, such as evaluating student guesses about scientific experiments, offering praise when appropriate, and prompting further elaboration if necessary, enhancing the educational interaction. Mollick [62] introduced seven AI personas for classroom use — mentor, tutor, coach, teammate, student, simulator, and tool — along with example prompts to assign specific roles to LLM chatbots. These examples highlight the initial exploration of AI personas in education. Further investigation of diverse roles and tailored applications of AI personas will be essential to unlocking their full capacity to enhance learning and teaching experiences.

2.4 Learning History from AI Chatbots

History education has a unique motivation to integrate AI chatbots compared to other subjects. Unlike math or science, where the focus is on solving problems or memorizing formulas, history emphasizes an affective engagement with historical facts and figures in order to better understand and contextualize past events, social issues, experiences, and actions [23]. Long before the advent of LLMs, early systems demonstrated the potential for conversations with historical figures such as Charles Darwin [85], Albert Einstein's ghost [57], and Richard Nixon [11]-which is often called

"virtual immortality" [8]. These interactions engage students with opportunities to explore multiple viewpoints and cultivate empathy as they gain insight into the thoughts and motivations of people from different eras.

Most of the early chatbots that demonstrate historical figures are hand-crafted or retrieval-based, based on existing material found in some text collections or knowledge graphs [73]. However, the development of historical chatbots faces new research opportunities to leverage the capabilities of large language models (LLMs). Such LLM-powered chatbots could go beyond simply delivering facts, instead engaging students in dynamic, immersive dialogues that reflect nuanced beliefs, cultural contexts, and the rhetoric of historical personalities. Such applications open avenues for personalized learning, foster critical thinking, and help students engage deeply with historical arguments and perspectives. Research could explore how these historical figure chatbots perform in teaching historical empathy [23] by simulating the emotions and worldviews of characters from different eras.

2.5 Historical Empathy

Historical empathy is a central factor in the construction of historical meaning and a powerful tool to understand history [41]. To gain historical empathy, students must engage both cognitively and affectively with historical figures to reconstruct, understand, and critically evaluate their lived experiences, decisions, and actions [9, 22]. The cognitive dimension involves contextualization, which requires the learner to reconstruct the values, beliefs, and motivations of historical actors based on historical evidence [19, 31]. On the other hand, the affective dimension emphasizes emotional engagement with historical figures, fostering resonance and moral reflection that motivate deeper inquiry [22, 30, 42].

As cognitive and affective dimensions are widely accepted, researchers have reinterpreted or expanded them further. For example, Endacott and Brooks [23] redefined the original cognitive dimension by distinguishing "contextual understanding" and "perspective taking" as two separate components. Karn [41] with her five-component model introduced four new components: evidence-based contextualization, informed historical imagination, perspective taking, and ethical judgment. Barton and Levstik [7] have also refined the cognitive dimension by introducing the concept of "perspective recognition." On the other hand, according to Barton and Levstik [7], the affective dimension is less about uncritical sympathy and more about developing a sense of "care" that motivates deeper inquiry into the past.

In this study, we adopt the Endacott and Brooks' model of historical empathy [23] to examine how AI-powered historical chatbots can support core dimensions of history learning. Specifically, we focus on four interrelated dimensions: perspective-taking, cognitive dimension (contextual understanding), affective dimension, and personalization, which together capture both established goals in history education and emerging needs in learner-centered AI design. We first conducted a formative study to explore how students and teachers perceive the potential and limitations of AI personas across these dimensions. Building on these insights, we developed HistoChat, a system that embeds these components into the design of AI historical figure agents, and conducted a follow-up user study to assess whether such a system can meaningfully support personalized and empathetic engagement with the past.

3 Formative Study

To explore the limitations of traditional history education and the potential of AI personas, we conducted a formative study with three teachers and eight middle school students. The study examined how students and teachers interacted with AI-powered historical figures, focusing on their expectations, usability, and perceived educational value. Insights from this formative study

informed the design of HistoChat, which was later explored in a main study to investigate its potential to support historical empathy and engagement.

In the following subsections, we describe the participant recruitment process and study procedure, including the structure of each phase and the study context. We then summarize the key findings that shaped the subsequent system design.

3.1 Participants

All participants were recruited in South Korea. Our participants consisted of three teachers (ages 21–23) and eight middle school students (three aged 13, five aged 15), as shown in Table 1. Teachers were selected based on their experience in teaching and interest in AI, reflecting perspectives of educators familiar with current curriculum requirements and open to experimenting with emerging educational technologies. All selected teachers had at least one year of small classroom teaching experience in history. For students, participants under 12 were excluded to reduce potential gaps in historical knowledge. To encourage natural and comfortable interaction, students were recruited in groups of peers who were already acquainted. Moreover, we included students with varying levels of interest and familiarity with AI to ensure diverse perspectives. We organized the teacher and student participants into three groups, each consisting of one teacher and two or three students.

Table 1. List of Participants for the Formative Study

Group	ID	Affiliation	Gender	Age	AI Interest	AI Familiarity
Group 1	T1	Teacher	Female	23	Very Interested	Somewhat Familiar
	F1	Student	Male	14	Very Interested	Very Familiar
	F2	Student	Male	15	Interested	Familiar
	F3	Student	Female	15	Interested	Unfamiliar
Group 2	T2	Teacher	Male	23	Interested	Familiar
	F4	Student	Female	13	Uninterested	Somewhat Familiar
	F5	Student	Female	13	Somewhat Interested	Somewhat Familiar
	F6	Student	Female	13	Uninterested	Unfamiliar
Group 3	T3	Teacher	Male	21	Interested	Familiar
	F7	Student	Female	14	Uninterested	Familiar
	F8	Student	Female	15	Uninterested	Very Familiar

3.2 Procedure

The formative study consisted of three phases: preliminary interviews with teachers, classroom observation, and follow-up interviews.

(1) **Phase 1: Preliminary Teacher Interviews.** We conducted individual Zoom interviews with teachers to explore the challenges of history education and their views on using AI-powered historical personas. Teachers described difficulties in engaging students and reflected on how conversations with historical figures could address these issues. Drawing on their curricular expertise, we worked with teachers to identify topics that were persona-centered, aligned with students' curriculum, and supported clear learning objectives. To capture a range of historical contexts and explore how different topics might shape interaction, each group was assigned a distinct topic: absolute monarchy in Europe (Group A), the Three Kingdoms of Korea (Group B), and the late Joseon Dynasty (Group C). Each teacher was asked to prepare a 30–35 minute lesson on their assigned topic, along with a set of three to five time-feasible questions related to the lesson content.

(2) **Phase 2: Classroom Observation and AI Interaction.** On the day of the study, each session began with self-introductions and a short activity where participants shared their favorite historical figures to create a comfortable atmosphere. Next, students configured their favorite AI personas in GPT-4 during a brief tutorial, using the prompt: "you are now [historical figure]. Please answer the questions I ask as if you are [historical figure]." Each teacher then delivered a 35-minute lesson on the assigned topic, after which students individually worked on a set of prepared questions. Throughout both the lesson and problem-solving, students were given the option to interact with the AI personas at any time or work without AI assistance. This approach was intended to prevent artificial constraints on students' learning processes and to allow authentic patterns of engagement with AI.
(3) **Phase 3: Follow-up Interviews.** The final phase included follow-up interviews in two parts. First, we conducted group interviews with students and teachers to discuss overall impressions, perceived strengths and weaknesses, unmet needs, and potential applications in real-life history learning. Then, teachers were interviewed individually to reflect on the class experience, anticipated challenges, and their perceived roles in integrating AI into their teaching practice.

All sessions were held in private group conference rooms in Korean, where each small-group lessons were conducted sequentially. Researchers and a teacher were present throughout to monitor activities and address any inaccurate or misleading AI responses. Informed consent was obtained from all participants and their guardians.

3.3 Analysis

With our goal to inform the design of HistoChat, we conducted a thematic analysis combining inductive open coding and theory-driven categorization. Two researchers independently reviewed the student–AI conversation logs, preliminary interviews, session recordings, and follow-up interviews. Following established qualitative procedures [17], we first generated initial codes through open coding to capture salient patterns in the data. These codes were iteratively discussed and refined to develop a shared codebook. We then grouped the codes into higher-order themes guided by the historical empathy framework and personalization. Rather than simply cataloging interaction preferences or usability issues, our goal was to understand how students and teachers interpreted the educational potential and limitations of AI-powered historical figures. Specifically, we examined the findings in relation to four interrelated educational dimensions: personalization, and the perspective-taking, cognitive, and affective dimensions of historical empathy. Our findings were then translated in English.

3.4 Formative Study Results

The analysis surfaced key insights across three areas: (1) restraints in traditional learning, (2) user interpretation of AI historical figures, and (3) observed hurdles in AI-based learning (see first three columns of Table 2). Teachers and students alike found that AI personas made history more interactive, personalized, and emotionally engaging by mitigating existing constraints. However, they also identified challenges that point to the need for more intentional design. To address these, we derived the fourth category: (4) design implications for AI-integrated history learning (see final column of Table 2). These implications from user expectations support designing HistoChat that help form personalized and empathetic connections with historical content. In the following sections, we report our findings across four dimensions—personalization, perspective-taking, cognitive, and affective. In addition to these four dimensions, we also report technical issues that arose during our formative studies for further considerations.

3.4.1 Personalization. Personalization is widely recognized as essential for promoting student engagement, yet traditional lecture-based classrooms often lack the flexibility to support individualized learning pathways [18, 92]. Teachers managing large groups have limited capacity to offer tailored attention, which students in our study frequently noted. Our participants recalled the past hesitance to ask questions due to constraints in time or teacher availability. As F3 put it, *"usually, if I wanted to look something like this up, it would be difficult to ask the teacher and get an answer during the short time."*

In contrast, AI historical figures enabled students to pursue self-directed inquiry. Students reported greater comfort asking questions and receiving immediate, personalized responses. F2 explained, *"there are often times when I don't fully understand something... here I was able to immediately look it up (through AI)."* Likewise, the teachers also recognized this benefit. T2 commented, *"students can get personalized explanations one-on-one, without needing to ask me for additional details."* Still, some students remained hesitant to use AI during our class, citing a perceived need to focus on the teacher—suggesting that classroom norms can inhibit autonomous interaction with AI.

Beyond teacher-centered barriers, the standardized, one-size-fits-all curriculum limits exploration tailored to students' varied interests and prior knowledge [18]. In our study, students used AI to extend learning beyond the lecture. F1 said, *"I was able to quickly learn things that aren't included in the textbook,"* and F3 added, *"it allowed me to expand my background knowledge more broadly."* Students expressed that they are *"especially interested in learning about the broader background of historical figures—like the situations they faced and how those shaped their actions,"* but *"it's not easy to find that kind of information in textbooks"*; however with AI historical figures, they *"could easily access those contextual details, which [they] really appreciated"* (F1, F2, F3). However, few students also found the AI's breadth and complexity were sometimes overwhelming. F2 noted, *"it can be difficult to understand all at once... I think it might be challenging to synthesize all the information."* Teachers echoed this concern, as T1 warned of *"bringing up vast and irrelevant content outside of curriculum."*

Together, these findings suggest two design implications: Empowering Autonomous Exploration and Stretching Students' Learning Boundaries. The former calls for minimizing teacher dependence by supporting student-led interaction with AI. The latter emphasizes scaffolded delivery of contextually appropriate information just beyond the textbook—challenging yet accessible. These directions aim to foster more adaptive and self-directed learning in history classrooms, combining freedom of inquiry with personalized support.

3.4.2 Perspective-Taking. In the historical empathy framework, perspective-taking refers to the cognitive effort to understand how people in the past thought, felt, and made decisions within their historical context [7]. Yet in traditional classrooms—relying heavily on lectures and textbooks—students often encounter history as a linear narrative, limiting their ability to consider diverse perspectives or engage with the reasoning behind historical actions. As one student noted, this restricts "*the ability to expand thinking beyond [textbooks]*" (F8), leading to passive absorption rather than reflective interpretation. Teachers likewise acknowledged that conventional methods tend to flatten historical figures into distant abstractions rather than multi-dimensional individuals situated in specific contexts.

In contrast, AI historical personas supported this dimension of empathy by enabling students to participate in first-person dialogue with historically grounded characters. These conversations allowed students to enter the mindset of historical figures, explore context-specific dilemmas, and reason through their motivations and constraints. The students described this interaction as immersive—*"I was curious about how my own situation or feelings might relate... I was able to get a*

Table 2. Summary of Results from the Formative Study

Dimension	(1) Restraints in Traditional Learning	(2) User Interpretation of AI Historical Figure	(3) Observed Hurdles of Utilizing AI Historical Figure	(4) Design Implications for AI Integrated History Learning
Personalization	Teacher-Centered; Limited Flexibility and Individualized Attention	AI Provides One-on-One Interaction; Supports Self-Paced Inquiry.	Teacher Authority Limits Students' Freedom to Use AI.	**Empowering Autonomous Exploration:** AI Should Minimize Teacher Reliance and Support Student-Led, Self-Paced Inquiry.
	Fixed Curriculum; Constrained Opportunities for Extended Exploration	AI Expands Access beyond Textbook Scope.	AI Responses Misaligned with Students' Depth or Needs.	**Stretching Learning Boundaries:** AI Should Provide Content Just beyond Current Understanding to Expand Knowledge.
Perspective Taking	Uniform Instruction; Limited Exploration of Historical Viewpoints	AI Enables Immersive, First-Person Conversations with Historical Figures.	Dialogue-Based Learning Is Unfamiliar; the Responses Feel Limited or Superficial[a].	**Anchoring Dialogue in Instructional Goals:** AI Should Sustain Perspective-Driven Dialogue Aligned with Learning Goals.
Cognitive Dimension	Static Textbooks; Limited Access to Diverse and Layered Perspectives	AI Offers Diverse Perspectives and a Nuanced Multi-Dimensional Understanding.	One-Sided Responses of AI Prevents Critical Thinking and Reasoning.	**Scaffolding Critical Historical Reasoning:** AI Should Offer Information Incrementally and Encourage Critical Interpretation.
			AI Reliance Discourages Inquiry and Active Thinking.	**Fostering Independent Historical Inquiry:** AI Should Spark Curiosity through Open-Ended Prompts and Motivate Active Exploration.
Affective Dimension	Difficulty Cultivating Internal Interest	AI Dialogue Evokes Interest and Lasting Impression with Emotional Resonance.	Gap in AI Prompting Skill Leads to Disparities in Emotional Connection and Depth of Learning.	**Self-Adapting AI Based on User Ability:** AI Should Adjust to Student Ability to Ensure Equitable and Engaging Interaction.
	Standardized Content; Difficulty Connecting with Personal Interests	AI Fosters Sympathy by Aligning Conversations with Personal Interests.	List-Style AI Responses Resemble Textbooks, Lacking Dynamics Expected from AI.	**Aligning Learning with Individual Passion:** AI Should Connect Content to Personal Interests to Increase Sympathetic Connection.

[a] The Finding in This Cell Is Expressed Only by Teachers. All Other Cells Reflect Both Student and Teacher Perspectives.

sense of [the historical figure's] feelings" (F6)—and emotionally engaging—*"it was fun because it felt like I was really talking to a historical figure from the past"* (F5).

However, teachers noted that learning through conversational dialogue can be more challenging compared to the clear and structured language of textbooks (T3), and can cause students to lose learning objectives due to feeling limited or superficial. Some students perceived the dialogue as overly casual or lacking in depth, leading to concerns that the experience could feel more like entertainment than meaningful learning. As T3 reflected, *"it might not really even feel like a class anymore."* These tensions suggest that while conversational interfaces offer promising pathways for perspective-taking, they also risk superficial engagement if not properly framed. Still, many students saw the dialogues as meaningful learning moments, appreciating both the novelty and enjoyment of engaging directly with historical figures.

To fully support this aspect of historical empathy, AI personas must go beyond role-playing to foster interpretive inquiry. This leads to a key design implication: Anchoring Dialogue in Instructional Goals. AI systems should structure conversations to support curricular aims and guide students toward deeper historical reasoning, ensuring perspective understanding is grounded in contextually meaningful reflection.

3.4.3 Cognitive Dimension. Within Endacott's framework of historical empathy, the cognitive dimension refers to students' ability to reason through the broader historical context—analyzing cause and effect, interpreting decisions within their structural constraints, and situating events in relation to broader patterns [23]. Unlike perspective-taking, which centers on entering the mindset of historical actors, the cognitive dimension emphasizes making sense of why things happened, based on contextual evidence and complexity. Traditional history education often limits students to surface-level understanding, offering fragmented facts without opportunities to synthesize broader patterns or critically evaluate complex events. For example, T3 mentions that *"it's important to be able to engage in multi-layered thinking, but this aspect is challenging [with traditional curriculum]. There should be a way to encourage more multi-dimensional thinking, but I found it difficult to set up such an approach."* This restricts critical thinking and prevents students from developing a multi-faceted view of historical events.

In contrast, AI-powered historical figures were found to offer a more dynamic approach to learning, allowing learners to access historical content from multiple angles and trace connections across contexts. This ability to deliver information from various angles helps students form a more comprehensive and multi-dimensional understanding of history, expressed by F1's comment *"when learning about history, we usually only see what is written about the figure in textbooks. But by using this, I could learn about what they were thinking, the effects of their actions, and the broader context surrounding them. It allowed me to see the information in a more comprehensive and multi-dimensional way rather than just seeing fragmented pieces, which I found really valuable."* This flexibility encourages students to think critically about the material and engage with it in ways traditional textbooks cannot.

However, there were two challenges observed. When AI simply delivers all information at once, it can diminish opportunities for critical thinking, making the interaction feel one-sided. As F2 noted, *"it tells you everything beyond what's in the textbook or worksheets... it feels like there's nothing left for you to think about on your own,"* while F1 added, *"since almost all the information is provided, relying solely on this can reduce opportunities to think independently."* Such encounters prevented further questioning from students, often ending the interaction prematurely. Without prompts that invite deeper inquiry, the students risk becoming passive recipients rather than active agents of inquiry. Users also raised concerns about overreliance on AI, which also lead to passive learning and discourage independent thinking. As T1 noted, *"in cases where it's misused, like when students don't want to focus on the lesson and just say random things,"* and T2 added, *"students might become too reliant, focusing only on what it provides."* Therefore, while AI has the potential to offer rich, multi-faceted insights, it is essential to encourage students to use the tool thoughtfully and actively, ensuring that it fosters critical thinking rather than passive consumption.

To mitigate these issues, we outline two complementary design implications: Scaffolding Critical Historical Reasoning and Fostering Independent Historical Inquiry. The first implication entails breaking down information delivery into stages that allow students to draw their own conclusions and interpret complexity incrementally. The second requires prompting students with open-ended questions and cues that spark curiosity and reflection. Together, these strategies aim to position AI personas not as content providers but as critical thinking partners.

3.4.4 Affective Dimension. Within the historical empathy framework, the affective dimension refers to students' capacity to emotionally engage with people from the past—recognizing their humanity, struggles, and moral complexity [7]. Unlike cognitive reasoning or perspective-taking, affective engagement centers on how students relate to historical figures on a personal and emotional level [23]. This includes feelings of curiosity, sympathy, admiration, or connection, which can deepen reflection and enhance retention. However, affective engagement is difficult to cultivate through conventional instruction, which often prioritizes factual delivery over emotional resonance.

A key challenge lies in cultivating learning motivation. As T3 noted, educators frequently struggle to *"[spark] that kind of primal curiosity that arises from within using external influences."* Students, however, described how emotionally charged dialogues with AI historical figures could evoke this internal drive. These interactions, often vivid and personally resonant, prompted students to reflect more deeply on their own aspirations. For example, F6—who had previously expressed strong dislike for history and a preference for science—reflected, *"I started thinking that I want to do something I like... [Jang Yeong-sil's] answer felt like it might have a significant impact on my future."* She added, *"I used to hate history the most, but now I think it might actually start to feel fun,"* noting that conversing with a historical figure known for scientific contributions sparked unexpected engagement. Yet, disparities in AI prompting skill introduced variability in experience: some students struggled to formulate effective questions, which led to less meaningful exchanges.

As S1 explained, *"people who are familiar with AI tend to ask questions effectively, but those who aren't used to it struggle...,"* and T1 warned that such differences could *"vary too much [and lead] to uneven outcomes."*

Beyond motivation, affective engagement also depends on students forming sympathetic connections with historical content. Traditional instruction, often standardized and lecture-driven, rarely connects content to students' interests. In contrast, AI personas could adapt to students' curiosities, helping them see historical figures as more relatable and emotionally real. F7 commented that the AI was *"giving me information that's more aligned with what I'm specifically interested,"* and T2 found that this alignment *"helped students stay focused and engaged."* However, when AI responses turned sequential or list-like—especially noted by F2 and F7—they began to resemble textbook delivery styles students hoped to move beyond.

These findings point to two key design directions for affective engagement: Self-Adapting AI Based on User Ability and Aligning Learning with Individual Passion. First, it is critical to ensure that all students—regardless of their prior AI experience—can access emotionally resonant learning. By adapting to individual prompting skills and cognitive capability, AI personas can promote more equitable affective learning outcomes. Second, the latter emphasizes the need for AI personas to connect with students' personal interests and emotional contexts. Doing so fosters sympathetic engagement and supports motivation that arises from within, rather than through extrinsic incentives.

3.4.5 Technical Issues. The risk of AI-generated misinformation—widely recognized in LLM research [5]—is especially consequential in history education, where accuracy and contextual nuance are essential. In our formative study, both students and teachers acknowledged this concern. One student noted, *"AI can give you wrong information"* (F8), while another reflected, *"users are quick to doubt AI, but still want to believe it"* (F5). This tension between skepticism and trust highlights a vulnerability in AI-mediated learning, where students may be persuaded by content they cannot fully verify. Teachers echoed this concern and suggested built-in referencing mechanisms to support source transparency (T1, T3). Yet, this absence of citation also prompted interesting engagement: F8 expressed curiosity about the primary source cited by the AI, prompting T3 to reflect that *"it was fascinating to see the student wondering where the material came from after hearing the AI's viewpoint,"* and view this as part of the learning opportunity to question the AI and reflect more deeply on the information. Still, teachers emphasized that their role must extend beyond managing student engagement or preventing over-reliance on AI; they are also responsible for regulating the flow of misinformation, ensuring that students do not internalize historically inaccurate content.

4 System Design

Based on the formative study results, we developed HistoChat[1], an LLM-powered conversational agent system that facilitates learning through interactions with the historical figure (hereafter, HF). To validate the design directions identified during the formative study, and to examine their impact on students' historical empathy, we implemented two system variants: the *Baseline* and the *Experimental* HistoChat.

Although both systems employ the same language model (GPT-4o) and a web-based user interface, they implement distinct prompting strategies, compared in Figure 2. In particular, the HF in the *Baseline* version passively responds to students' queries. Conversely, the *Experimental* version is designed to foster students' historical empathy by actively engaging them with personalized questions, adapting to each student's interest, concerns, and level of knowledge in history. The

[1] https://HistoChat-bbf8e.web.app/

formative study findings, associated design goals, and actual prompts utilized within both *Baseline* and *Experimental* versions are elaborated in the following sections.

4.1 Prompt Engineering

To establish the basic setup for conversations, both *Baseline* and *Experimental* systems used three basic prompts (1-3 in Figure 2c). **Prompt 1** set the [HF], instructing the system to respond as if it were the designated figure, adopting the tone and language appropriate to the era of the figure. **Prompt 2** defined the learning context, specifying that the conversation would be with a middle school student and that the goal was to teach a particular topic through dialogue. It also encouraged the system to actively guide the conversation toward the learning objective and to ask engaging questions. Lastly, **prompt 3** focused on maintaining continuity by referencing the previous dialogue, ensuring that the system was built on earlier exchanges while moving the conversation in a new direction, and avoiding repetition of previously discussed information.

Unlike the *Baseline*, the *Experimental* one features experimental prompts (4-7 in Figure 2d) derived from the findings of the formative study, in addition to the basic prompts. First, **prompt 4** provides the students with three anecdotes experienced by the [HF] and allows them to choose one. Instructed by **prompt 5**, the figure gradually gains a deeper understanding of the students' situation. We designed prompts 4 and 5 to tailor the conversation to each student's life and to create and maintain a deeper sense of relevance and personal connection throughout the conversation. In addition, the prompts allowed HistoChat to adapt to the students' diverse levels of familiarity with AI so that they can engage meaningfully in emotionally rich conversations with the [HF].

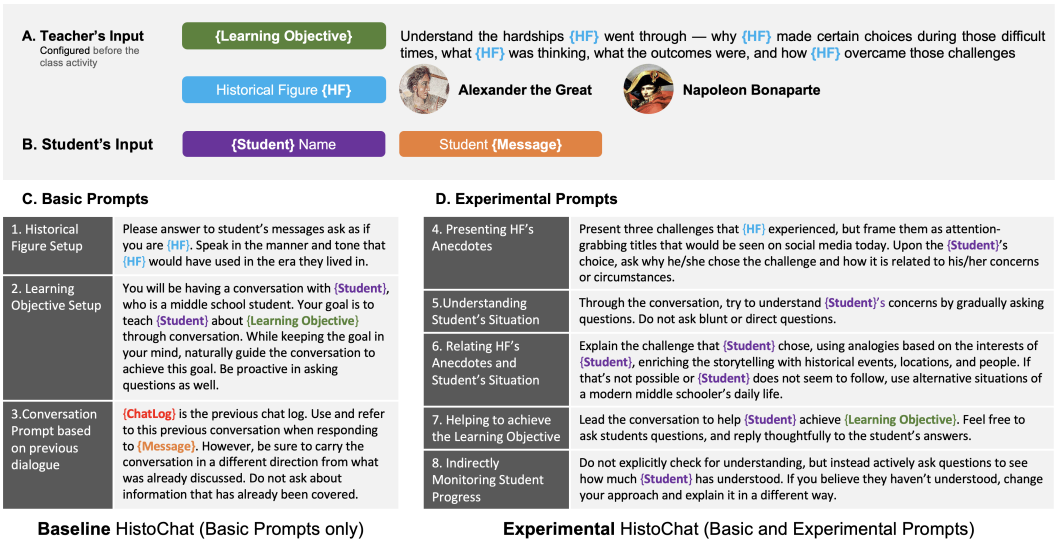

Fig. 2. An Overview of HistoChat Prompt Structure. Prior to the classroom activity, (A) the teacher sets a learning objective and HFs. At the beginning of the session, (B) students type in their names. Both the *Baseline* and *Experimental* conditions use the basic prompts (C) that instruct (1) how to use the HFs, (2) the learning objective, and (3) to utilize prior dialogue. In addition to the basic prompts, the *Experimental* condition employs experimental prompts (D) that comprise (4) presenting three anecdotes that the HF experienced and relevant to the learning objective, (5) understanding the student's situation, (6) relating the chosen anecdote and the student's situation, (7) leading the conversation to achieve the learning objective, and (8) checking the progress.

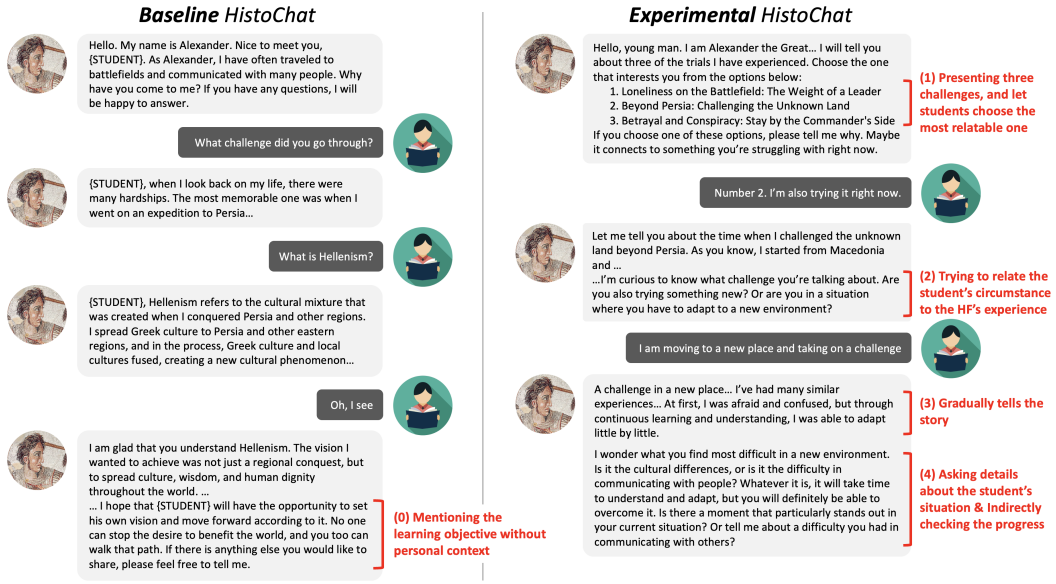

Fig. 3. An Example Conversation in Both Versions of HistoChat. (Left) The HF in the *Baseline* passively responds to the student's questions, trying to address the learning objective when possible. (Right) The HF in the *Experimental* version (1) begins with three challenges, (2) tries to relate the student's circumstance to the HF's experience, (3) gradually tells the story, and (4) asks for more details about the student's situation and indirectly checks the student's progress.

Subsequently, **prompt 6** connects the anecdotes to the student's situation by gradually unfolding the story through analogies and enriching the story with historical context (e.g., dates, locations, notable figures, and key events). This method was aligned with the design objective of scaffolding critical reasoning by breaking complex information into smaller parts so that students can progressively develop their own interpretations without experiencing steep learning curves.

Prompt 7 fosters mixed-initiative communication between the [HF] and students. Although most educational chatbots, as well as the *Baseline* version, focus on answering questions, the *experimental* system was crafted to stimulate curiosity, critical thinking, and self-directed exploration of historical content through open-ended questions and reflective cues.

Lastly, **prompt 8** makes the [HF] indirectly checks the student's progress toward the learning objective by asking questions, and adapts its teaching strategy as needed. A complete list of prompts can be found in the Appendix A.

4.2 An Example Scenario of *Baseline* and *Experimental* HistoChat

This section compares how the two versions affect students' experience. Before the classroom activity with HistoChat, the teacher needs to set the learning objective and HF. In this walkthrough we assume that the teacher has set Alexander the Great as the HF, and the learning objective as follows.

> Understand the hardships Alexander the Great went through — why he made certain choices during those difficult times, what he was thinking, what the outcomes were, and how he overcame those challenges.

Once the student joins the chatroom, *Baseline* greets with a short introduction, saying *"as Alexander, I have often traveled to battlefields."* Later on, the *Baseline* version passively responds to the student's messages, although often tries to remind the learning objective.

In contrast, the *Experimental* version begins with three anecdotes that the HF experienced and are relevant to the learning objective provided. Upon the student's choice, the HF gradually tells the story while skillfully connecting with the student's situation. For example, Alexander the Great in Figure 3 begins by framing his campaign beyond Persia, emphasizing that it was not just about military conquest, but also about *facing the unknown*. He then openly admitted feeling afraid and confused when faced with a new culture, language, and customs, saying *"at first, I was afraid and confused, but through continuous learning and understanding, I was able to adapt little by little."* This vulnerability mirrors the emotional challenges many students face when moving to a new place or school. On the other hand, Alexander did not just give a lecture, but he turned the conversation back to the student, asking *"are you also trying something new?"* Such open-ended, reflective questions are to strengthen historical empathy in the student's mind. Lastly, Alexander reflected on how collaboration and respect for local culture helped him overcome the challenges. He then transferred the lesson to the student, saying *"it will take time to understand and adapt, but you will definitely be able to overcome it."*

In sum, the conversation with Alexander demonstrates that the *Experimental* version enables the HF to build historical empathy and engagement by showing that even legendary figures faced similar challenges, and the lessons they learned are applicable to the student's circumstances. We believe that the *Experimental* version not only teaches historical facts, but also makes history more emotionally engaging, reflective, and perspective taking, which are the core design goals identified in the formative study.

5 Main Study

We conducted user studies to explore how the *Baseline* and *Experimental* versions of HistoChat shaped the learning experiences of middle school students, with a focus on emotional connection, reflective thinking, and perspective-taking behaviors. Rather than focusing on evaluating which system was more effective, our aim was to use these designs as complementary lenses to explore a richer set of student experiences and deepen understanding of how learners engage with AI-powered historical personas.

5.1 Participants

Twenty-six middle school students (ages 12–15; 10 male, 16 female) and three history teachers (ages 22, 28, and 32; all male) were recruited in South Korea through local schools, parent internet communities, and academic institutes. Data from 25 students were included in the analysis, as one student had to leave during the study. Most of the participants reported being interested in history: 9 were *very interested*, 7 *moderately interested*, and 6 *extremely interested*, while only 3 were *slightly interested* ($n = 1$) or *not interested* ($n = 2$). Their familiarity with AI varied, with most describing themselves as *moderately familiar* ($n = 9$) or *somewhat familiar* ($n = 7$), followed by *slightly familiar* ($n = 4$), *very familiar* ($n = 4$), and *not familiar* ($n = 1$). None had previously used AI in an educational context. In contrast, the participating teachers reported being either *moderately familiar* ($n = 1$) or *very familiar* ($n = 2$) with AI. Each student received a 20,000 KRW gift card (approximately 16 USD), and each teacher received 200,000 KRW (approximately 160 USD) in compensation. The participants were divided into three groups, each consisting of one teacher and eight to nine students.

5.2 Study Procedure

As outlined in Figure 4, the study adopted a within-subject design with three groups (Groups X, Y, and Z). Each group's sessions were held on different dates in Korean, lasting approximately 90 minutes as special one-time lectures conducted in a conference room at a local university. Each student interacted with two LLM-based historical figures - Napoleon Bonaparte in the first stage and Alexander the Great in the second. These two historical figures were selected in consultation with teachers, as they were considered appropriate for the middle school curriculum in Korea.

Both stages followed the same structure: a 5-minute teacher-led introduction, a 20-minute AI interaction, and a 10-minute post-test and evaluation, with a 10-minute break between stages. The sessions concluded with a cross-condition reflection and a brief teacher interview. The procedure was designed to promote an immersive learning experience and support teacher facilitation while closely simulating a classroom environment. To control for order effects, the sequence of experiencing the *baseline* and *experimental* systems was counterbalanced across groups. Neither students nor teachers were informed about the differences between the two systems, and the historical figures were simply referred to as 'AI Napoleon' and 'AI Alexander'.

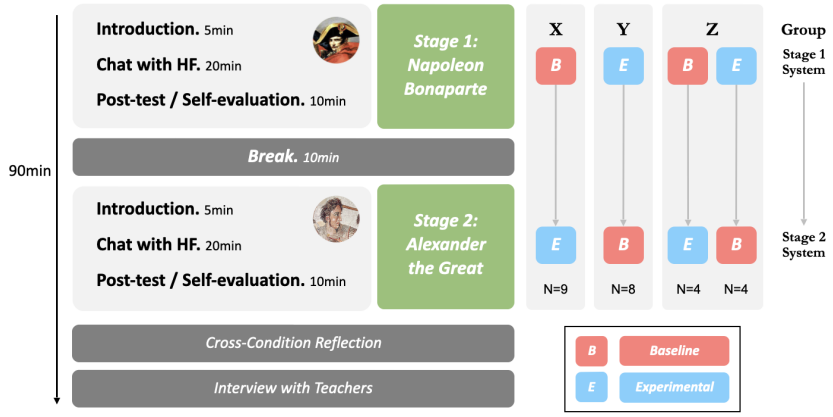

Fig. 4. An Overview of HistoChat Study Procedure. The within-subject study design followed a two-stage procedure involving both the *Baseline* and *Experimental* versions of HistoChat. A total of 25 middle school students participated in the study, organized into three group sessions (X, Y, and Z) with counterbalanced condition orders.

(1) **Introduction and Pre-Tests.** First, as an ice-breaker, students created name cards and introduced themselves, similar to the first day of class in schools, to help them feel comfortable in the classroom environment. The study began with a 5-minute introduction by the researcher, who provided an overview of the class and a tutorial on how to interact with the historical figures, which were framed as first-person representations. The system was designed to greet students in character upon login and initiate the conversation with a historically grounded voice. Students then completed a pre-test on Napoleon Bonaparte and Alexander the Great to assess prior knowledge. The pre-tests included five multiple-choice questions about historical facts and seven self-reflective questions measuring historical empathy (see Appendix B). The students were informed that their responses would remain confidential to reduce pressure and encourage honest reflection.

(2) **Stage 1: Napoleon.** The first stage focused on Napoleon and lasted 35 minutes. The history teacher began with a 5-minute introduction on Napoleon, providing minimal background

knowledge for students to converse with AI chatbots. The students then engaged in a 20-minute one-on-one dialogue with the AI Napoleon with the learning objective: *"understand the challenges that Napoleon faced and how he overcame them,"* and completed worksheets based on this objective, including questions such as: *"what challenges did Napoleon face? What choices did he have and why? What wisdom did he share? What would you have done?"* After interacting with the AI, students completed a closed-book post-test and an evaluation survey of the HistoChat experience. The evaluation included 20 5-point Likert scale questions (10 questions on historical empathy and 10 on personalization; see Appendix C) and 4 open-ended questions to understand their experience with the AI system.
The four open-ended questions were: "how was your history lesson with <AI Napoleon>?", "what impressed you during your interactions with <AI Napoleon>?", "what were the downsides of <AI Napoleon>?", and "what do you think AI is after interacting with a <AI Napoleon>?" After the evaluation, the students received a 10-minute break before going into the second stage where they used the alternate AI system embodying Alexander the Great.

(3) **Stage 2: Alexander the Great.** After the break, the students began the second 35-minute stage, which followed an identical structure but featured Alexander the Great as the embodiment. As in the first stage, students completed a post-test and an evaluation of their AI experience at the end of the stage.
(4) **Cross-Condition Reflection.** After completing both stages, the students compared their experiences with the two systems focusing on interaction style, emotional engagement, and perceived usefulness (see Appendix D).
(5) **Interview with Teachers.** A 30-minute 1:1 interview was conducted only with the teachers to gather additional insights into student engagement and their views on how AI systems can be integrated with history education.

To ensure consistency across stages and across different groups, researchers provided detailed class guidelines to teachers prior to the study, and checked teachers' classroom material. Following these guidelines, the background content delivered by teachers remained consistent across sessions and among all three class groups, as established by prior researcher instructions. Additionally, to minimize bias across two systems, teachers were instructed to maintain strict neutrality throughout both stages, using consistent language, tone, and refraining from suggesting prompts or intervening in AI-student interactions.

5.3 Data Analysis

We adopted a mixed-methods approach to analyze how students engaged with the two AI persona systems. Quantitative data included pre- and post-tests assessing content knowledge and evaluation surveys measuring perceived engagement, personalization, and historical empathy. Paired-sample t-tests were conducted to assess changes in knowledge between sessions, while descriptive statistics were used to compare engagement and empathy ratings across conditions.

Qualitative data included open-ended survey responses, student worksheets completed during the AI interaction, post-session reflection activities, and transcripts from one-on-one teacher interviews. Thematic analysis [25] was conducted following Braun and Clarke's six-phase method [15]: familiarization, initial coding, theme identification, theme review, definition, and reporting. Two researchers independently coded the data, followed by iterative discussion to resolve discrepancies and refine the theme structure. Together, these analyses provided an integrated view of how AI historical figures shaped students' emotional and cognitive engagement with historical personas and its potential in future history education. All qualitative data were analyzed in Korean and then formally translated into English for reporting.

5.4 Ethical Considerations

All procedures were reviewed and approved by the IRB board. Written consent was obtained from students and their parents or guardians prior to participation. To ensure student safety and minimize potential harm, a teacher and a researcher were present during all sessions to monitor AI interactions in real time. In cases where the AI provided potentially inappropriate responses, teachers were permitted to clarify or correct information as needed.

6 Results

Across both *Baseline* and *Experimental* HistoChat, we quantitatively analyzed data from 25 students, focusing on measures of interest, familiarity, and session-specific interaction dynamics. The survey results indicated moderate to positive engagement with both history and AI. The students reported a moderate *interest in history* ($M = 3.64, SD = 1.15$), with similar scores for *liking* ($M = 3.60, SD = 1.19$) and slightly lower levels of *understanding* ($M = 3.40, SD = 1.00$). *Interest in AI* was also moderate ($M = 3.64$, $SD = 1.08$), with a median of 4.0 and a 75th percentile of 5.0, indicating a strong interest in some students. *Experience with AI* ($M = 3.44, SD = 1.04$) and *familiarity* ($M = 3.48, SD = 1.00$) suggest foundational awareness.

For each stage of AI interactions (approximately 20-minutes), participants made an average of 15.86 prompts ($SD = 9.01$). Each prompt contained an average of 301.44 tokens ($SD = 205.71$), and the historical figure responded with an average of 8,479.94 tokens ($SD = 4,402.34$). Tokens represent the basic unit of text that language models use to process and generate text. For GPT-4o, approximately one token equals about four characters in English (1–2 characters in Korean), with an average ratio of 1 word to 1.33 tokens. Based on these estimates, the participant prompts contained approximately 226 words on average, while the historical figure responses contained approximately 6,365 words [67]. A two-way ANOVA revealed that there were no statistically significant differences between the systems or stages.

The responses to these questions were quantitatively analyzed using paired-samples t-tests. The results of the paired-samples t-tests can be found in Table 3. Out of five questions on knowledge, one item (Q1) showed a statistically significant improvement in the *Baseline*, compared to two in the *Experimental* (Q1 and Q5), the results for the Self Reflective questions (Q6–Q12) showed mostly significant improvements. For both *Baseline* and *Experimental*, all items except Q12 demonstrated statistically significant differences between before and after the interaction. Given that the study setting involved minimal teacher interaction and primarily focused on AI-based conversations, the observed increases in self-perceived connection suggest that persona-based dialogue offers potential for supporting emotional connection with historical figures.

7 Findings

The results from the user study revealed key benefits of using HistoChat for learning history. First, students showed small differences in how they perceived AI after interacting with the two systems. Second, we focus on the broader characteristics of HistoChat in the context of historical education, highlighting both strengths and limitations. Lastly, we draw on how the interactive nature of the AI provided an engaging and dynamic way to learn historical content and build historical empathy. In the following subsections, we summarize these insights as follows: (1) the influence of each AI persona on students' perceptions of AI, (2) the qualities and drawbacks of AI-powered historical education, and (3) the benefits of engaging with AI chatbots as a mode of learning. Together, these findings help clarify the role of AI in supporting historical education and its potential limitations.

Table 3. Comparison of Paired-Samples *T*-Test Results between *Baseline* and *Experimental* Conditions

Q#	Question	*t* (Base)	*p*	*t* (Exp)	*p*
Knowledge Test					
Q1	Identification of Nationality	-2.87	**.008**	-2.45	**.022**
Q2	Reforms and Military Achievements	-1.55	*.13*	-1.23	*.22*
Q3	Military Leadership and Imperial Legacy	-1.12	*.27*	-0.98	*.33*
Q4	Military Decline and Unfulfilled Conquests	-1.97	*.06*	-1.78	*.08*
Q5	Governance and Mentorship in Historical Context	-1.77	.45	-2.32	**.029**
Self Reflection					
Q6	Life and Achievements	-7.86	**<.001**	-5.59	**<.001**
Q7	Battles and Outcomes	-7.62	**<.001**	-6.83	**<.001**
Q8	Thoughts and Perspectives	-7.41	**<.001**	-9.75	**<.001**
Q9	Leadership and Political Strategies	-5.63	**<.001**	-6.83	**<.001**
Q10	Modern Relevance	-8.06	**<.001**	-6.96	**<.001**
Q11	Personal Relevance	-4.00	**<.001**	-4.55	**<.001**
Q12	Influence on Future Actions	-1.21	*.24*	-1.08	*.29*

7.1 Influence of Each AI Persona on Student AI Perceptions

Prior research underscores the significance of user perception in determining how effectively AI systems are utilized, as it influences engagement, interaction, and trust in the technology [51]. Therefore, understanding how students perceive HistoChat offers valuable insight into their immediate reactions and interpretations of the technology. While these perceptions may not directly indicate whether the tools improve learning outcomes, they reveal how students conceptualize AI's role in their educational experience.

To investigate these perceptions, we asked students to respond to the question: " what do you think AI is, after your experience with historical figure?" after using each system. These reflections offered a snapshot of how students interpreted the nature and function of AI immediately after engaging in full-session dialogues with historically grounded characters. While the exposure to AI was relatively brief, and thus unlikely to yield deeply formed perceptions, these first impressions nonetheless offer meaningful insight into how students make sense of conversational AI in educational contexts. As a form of firsthand perception, they capture how AI is interpreted at the point of use—before extended reflection or normalization might flatten such responses.

By analyzing the students' responses, we identified two major dimensions of perception: (1) understanding AI and (2) understanding the impact of AI. In categorizing these perceptions, we drew on existing frameworks of AI competency for middle school learners to guide our classification [51]. Within these overarching categories, we further observed distinct tendencies in students' views across several subcategories. The resulting categories, illustrated with representative quotes, are summarized in Table 4. This table also indicates the number of students who expressed each perception, providing an overview of the distribution of perspectives across the sample. Insights derived from these observations are discussed in detail in the following subsections.

Through examining these perceptions, we can better understand how students approach and interact with AI systems—whether they see the technology as an extension of their learning or as something more autonomous or abstract. We focus on understanding these perceptions as they are, shedding light on how two systems shape students' views of AI's capabilities.

7.1.1 Understanding AI. When we analyzed students' broader perceptions of AI, we identified a subset who expressed views relating to understanding of AI's fundamental capabilities. Here, the

Table 4. Categories of AI perception from Students with Student Quotes

Categories of Students' AI Perceptions	*Student Responses from Baseline*	*Student Responses from Experimental*
Understanding AI		
Understanding AI as Its Functionality		
AI Works Based on Existing Data.	"AI Speaks Based on Alexander's Data." (S6) "A Machine that Knows Things Humans Have Ever Discovered." (S4)	
AI Can Think by Itself.	"Beings that Can Think and Answer for Themselves." (S14)	"Thinkable Beings." (S14) "[AI] Thinks for Itself and Puts Things Together." (S6)
AI Can Improve on Its Own.		"AI Can Tell You without Being Taught." (S3) "[AI Can] Further Develop and Improve." (S7)
AI Is Boundless.		"[AI Is] a Jack of All Trades." (S22) "A Machine that Can Do Almost Anything." (S4) "Technology that Is Boundless." (S15)
Understanding AI as a Conversation Partner		
AI Can Answer My Questions.	"[AI Can] Give Multiple Answers." (S3)	"An Instrument that's Good at Giving Me the Answers I Want." (S16) "[AI Can] Quickly and Easily Provides Answers to My Questions." (S13)
AI Can Interact in a Conversation.	"AI Is Capable of Conversing with Humans." (S23) "Able to Communicate with Me." (S9) "Being Able to Have a Realistic Conversation." (S12)	"[AI Is] Capable of Communication." (S9)
AI Listens to Me Attentively.		"[AI] Listened to Me Well and Responded Properly." (S17) "A Being that Listens to Me Well when I'm Having a Hard Time." (S23)
AI Is a Way to Connect Past with Present.		"[AI Is] a Way for the Past and Present to Communicate." (S10)
Understanding of AI Persona		
AI Is a Computer with Feelings.	"AI with Emotions." (S10) "A Smart Emotional Computer." (S26)	
AI Can Represent Someone.	"Not a Complete Copy of the Person, but Knows Their Thoughts." (S11) "[AI Is] a ChatGPT Representing Napoleon." (S22)	
AI Resembles a Real Person.	"A Machine that Feels Like a Real Person." (S20) "[AI] Feels Like a Real Person." (S5) "A Program Resembling a Real Person." (S23)	"It Feels Like a Human." (S5) "A Machine that Feels Like a Real Person." (S20) "[AI Is] a Program that Feels Real." (S12)
AI Is a Real Person.		"[AI Can] Replace a Part of Someone." (S1) "[AI Is] a Real Person, Not Virtual." (S11)
AI Is My Soulmate.		"[AI and I] Are Perfect Peas in a Pod." (S24)
Understanding Impact of AI		
In Education		
AI Is Knowledgeable.	"[AI] Knows More than I Do." (S24) "An Intelligent Machine." (S20) "AI Is a Smart Object." (S2) "AI Is Smarter than I Thought." (S17)	
AI Can Teach Me.	"[AI] Can Communicate with Me and Teach Well." (S9) "AI Is Convenient for Helping with Human Learning." (S8) "A Tool for Self-Study." (S25) "Something that Explains Various Ideas Well to Me." (S13)	"A Learning Tool that Can Greatly Assist My Self-Study." (S25) "An Entity that Can Communicate and Help Me Learn through Interaction." (S9)
AI Makes Learning More Captivating.		"[AI] Can Make Traditional Textbook Lessons More Enjoyable." (S25) "[AI] Allows for Fun, Engaging Learning Approaches so that It's Not Boring." (S2)
In Life		
AI Can Interest People.	"A Very Interesting Tool." (S15) "A Device that Makes Me Even More Interested." (S16)	
AI Can Help People.	"[AI] Is Something that Helps Me." (S1) "[AI] Is Helpful." (S21) "[AI] Is Helpful to Humans." (S7)	
AI Can Make Life More Convenient.	"[AI] Can Conveniently Resolve What I Have in Mind." (S18)	"AI Makes Convenient." (S2) "AI Is Conveniently Present Anytime and Anywhere for Help." (S18)
AI Improves Quality of Life.		"[AI Is] Enhances Our Quality of Life and Supports Us." (S8)

students perceived AI as a functional system, as a conversation partner, and as an AI persona. The responses suggest varying levels of understanding the AI, shaped by the system they used.

Understanding AI as a Functional System Students using *Baseline* generally viewed AI in traditional terms—as a tool that retrieves and processes existing information. This perception reflects a structured, data-driven understanding, where AI is bound by the knowledge it has already accumulated. In contrast, students using *Experimental* appeared to adopt a broader perspective, imagining AI as a more autonomous, self-improving entity. There was a notable openness to the idea that AI could think independently and develop over time, indicating a more dynamic and fluid interpretation of AI's capabilities. Additionally, several students using *Experimental* viewed AI as having virtually no limits, perceiving it as a versatile and all-encompassing tool. This contrast suggests that *Experimental* may have encouraged a more expansive understanding of AI's potential, while *Baseline* reinforced a more grounded and controlled view.

Understanding AI as a Conversational Partner Among students who perceived AI as a conversational partner, those using *Baseline* focused on its ability to answer questions efficiently. They often viewed AI as a dependable tool for retrieving information, with interactions framed as direct and goal-oriented. In contrast, students using *Experimental* emphasized the system's responsiveness and conversational quality. They described the AI as more attentive and thoughtful, creating a sense of dialogue rather than simple information retrieval. The *Experimental* version was often seen not only as a provider of answers, but as an engaged partner capable of sustaining meaningful exchanges. One student (S10) even described the *Experimental* version as a bridge between past and present, suggesting it could connect historical perspectives through dialogue. While the interactions were brief, students' differing responses suggest that perceived conversational quality is not just a matter of tone or output length, but how agency is distributed within the exchange. Even subtle differences in prompting structure may influence whether students treat the AI as a transactional tool or a dialogic presence—shaping how they engage with both the content and the character.

Understanding of AI Persona Many students perceived both versions as a persona, though they varied in whether they saw it as a mechanical tool or a more human-like entity. In *Baseline*, students often regarded AI as functional—able to simulate human traits but ultimately grounded in its role as a machine. While some acknowledged that *Baseline* could mimic thoughts or emotions (S10, S16), they primarily saw it as a representation rather than a true embodiment of a person. For instance, students noted that while the *Baseline* could replicate a historical figure's thinking, it was still clearly computational, not human (S11, S22). Still, across both systems, some students described the AI as feeling human-like, citing responses and behaviors they would expect from real people. This perception was especially prominent in *Experimental*, where a few students went so far as to say AI could replace part of a person or was "a real person" (S11). One student even described herself and the AI as *"peas in a pod"* (S24), revealing a view of AI as not just lifelike, but a deeply personal companion. For this student, AI blurred the line between simulation and genuine social connection. These moments suggest that even brief encounters with AI personas can prompt students to project relational and human-like qualities onto the system. While these perceptions may not indicate deep, long-term bonds, they reveal the potential for AI to be experienced not only as a learning tool but as a socially meaningful presence—particularly when designed with responsiveness and consistency in voice. Such reactions invite further exploration into how affective resonance and perceived companionship might influence students' sustained engagement with educational AI personas over time.

7.1.2 Understanding Impact of AI. Other students perceived AI on its potential impact within educational contexts or in their everyday lives. These reflections provided insights into how students viewed the practical applications and long-term potential of AI.

Impact of AI in Education In *Baseline* HistoChat, students frequently described AI as a knowledgeable resource, often referring to it as an intelligent machine or a *"smart object"* (S2) that knew more than they did. AI was seen as capable of providing valuable information, and appreciated AI's ability to explain ideas clearly, but it was largely viewed as a reliable tool for retrieving information rather than a highly interactive teacher. When considering AI's role as a teaching tool, students across both *Baseline* and *Experimental* acknowledged its ability to support their learning. AI was perceived as helpful for self-study, offering clear explanations and assisting in understanding complex ideas. While *Experimental* users highlighted a slightly more interactive, conversational teaching approach, students in both systems appreciated AI's capacity to guide their learning process effectively, with subtle differences in how engaged they felt in these interactions. Additionally, four students expressed AI to make the learning process engaging and enjoyable

after using *Experimental*. They noted that AI made traditional lessons more fun and captivating, offering engaging learning approaches that kept the material from feeling boring. This contrasted with *Baseline*, where the student perception in the interaction was more functional and focused on obtaining information efficiently.

The variation in prompting styles appeared to influence how students positioned AI within the learning process. While some viewed it primarily as a tool for information retrieval, others engaged with it as a more responsive instructional guide. These differing framings suggest that AI does not occupy a fixed pedagogical role, but rather that its perceived function is actively shaped by how it communicates. Even in short interactions, students seemed to interpret these patterns as cues—about whether AI was there to support, to lead, or to interact. Such framing has implications not only for engagement, but also for the kinds of expectations students form around AI's place in their educational experience.

Impact of AI in Life Some students reflected on AI's relevance beyond the classroom, highlighting how it could serve as a useful tool in their daily lives. In the *Baseline* condition, several students described AI as interesting, engaging, or generally helpful—suggesting a view of AI as a supportive utility. Across both systems, a few students mentioned that AI made life feel more convenient, emphasizing its availability and responsiveness as a key benefit. This practical framing positioned AI as a tool that could reliably assist with various everyday tasks. In the *Experimental* condition, one student extended this idea further, describing AI as something that enhanced their overall quality of life—a more integrated and broader perspective. We saw students naturally extended their perceptions to everyday life; they did not rigidly separate classroom learning from daily experience, but instead viewed AI personas as continuous with their broader ways of thinking and living. Given that middle school students are learning how to connect school-based knowledge with real-world relevance, it is perhaps unsurprising that they interpreted the AI not only as a learning assistant, but as a companion that could support personal goals or life direction. This tendency to associate educational AI with everyday purpose may also reflect a deeper daily connection with the historical figures—one that aligns with the affective dimension of historical empathy, where the past becomes meaningful in relation to the self.

7.2 Qualities and Drawbacks of AI-powered Chatbot in Historical Education

The main study revealed how the expectations from formative study played out in practice—surfacing not only benefits but also unforeseen challenges that arose during actual usage with two prompting strategies. To understand how HistoChat shape historical learning, we examined students' reflections on their interactions with HistoChat. Although not all components directly map onto historical empathy, many revealed learning dynamics that closely relate to its cognitive, affective, and perspectival dimensions [23]. Student responses on memorable aspects of each HistoChat are summarized in Table 5. While both systems were noted for the historical figures' personalities, students using the *Baseline* version more often recalled factual content as memorable, whereas those using the *Experimental* version more frequently highlighted their sense of personal connection with the persona. Then, specific qualities and drawbacks observed from each system are summarized in Table 6, which provides an overview of the findings discussed in this section and described in more detail in the following subsections.

The overview of the five findings explained in this section and flowchart of using HistoChat is presented as a teaser Fig.1, illustrating at a glance how the system design connects to the observed learning experiences.

7.2.1 Taking Initiative in Learning. In the formative study, students anticipated that AI might help them ask questions they would hesitate to pose to a teacher. In the main study, this expectation was

Table 5. Memorable Aspects of the AI Historical Figure

# of students	*B*	*E*
Facts about Historical Figure		
Core Beliefs and Principles	5	3
Life Journey and Legacy	5	4
Accomplishments	5	2
Passions and Personal Interests	1	2
Personality of Historical Figure		
Building Intimacy	1	2
Life-Like Presence (Human-Likeness)	5	4
Personal Relevance with Historical Figure		
Giving Me Advice	4	4
Associating with Me	2	7
Understanding My World	4	4

realized: students freely initiated questions and felt empowered to explore topics without relying on teacher guidance. This sense of autonomy emerged in three areas: learning independently, asking unrestricted questions, and as a result, feeling motivated to continue learning on their own.

First, students noted that AI explained concepts clearly without the restriction of a human teacher or a classroom environment. As they shared, *"AI taught me in a way that was easier to understand without my teacher"* by *"fully explaining in fun words"* (S8) and *"[explaining] new facts [they] didn't know in a clear and natural way"* (S4). Second, the freedom to ask any question was especially valued: as S19 put it, *"I was able to ask everything I was curious about."* This sense of control was more pronounced in the *Baseline* condition, where one student described the value of *"taking control of the conversation"* (S15). Third, the AI interactions sparked a deeper desire to continue learning. Several students reported that the experience motivated them to pursue further study on their own. As S7 and S14 shared, *"I wanted to know more about the history,"* and *"it came to my mind to study more history in the future,"* respectively.

However, there were some shortcomings in the sense of agency in the *Experimental* condition. The AI's tendency to take on the role of the teacher and ask directed questions limited students' freedom, in contrast to the *Baseline* condition. Following the AI's predefined scenario made the learning process feel more rigid, as they could not fully explore topics on their own terms when *"[Experimental] continued to steer the story"* (S21).

7.2.2 Receiving Effective and Engaging Answers. In the formative study, students anticipated that AI could deliver *useful answers* to their questions. In the main study, students consistently highlighted how the AI gave them the type of answer they sought—concise clarifications, extended explanations, or lively stories. Building on the earlier finding that students valued the freedom to ask questions without restraint, another benefit we found was that their satisfaction extended to the answers themselves. The emphasis here is less on simply being able to ask and more on receiving the desired response efficiently and enjoyably.

Students' satisfaction was not only about getting answers, but about how those answers matched their immediate learning goals. First, concise answers provided a sense of direct satisfaction, and these were especially frequent in the Baseline condition. As S12 noted, *"The best part was being able to pick only what I was curious about and get clear answers."* For moments of deeper inquiry, the AI extended explanations that broadened knowledge. By contrast, the *Experimental* version often captivated students with lively narratives, making history memorable and entertaining; as S2 shared: *"The story of Alexander was even more interesting and fun."* Taken together, these experiences reveal a spectrum of expectations—from efficiency to immersion.

Table 6. Qualities and Drawbacks of HistoChat, with Frequency of Student Mentions

# of students	*B*	*E*
Taking Initiative in Learning		
Being Able to Learn without a Teacher	1	2
Freely Asking Questions	7	4
Developing a Deeper Desire to Keep Learning	6	5
****Drawbacks***		
Continuous Questioning Limits Freedom, Reducing Flexibility	0	8
Receiving Effective and Engaging Answers		
Receiving Concise and to-the-Point Responses	8	2
Receiving Thorough and Comprehensive Responses	2	4
Enjoying Engaging and Entertaining Responses	4	7
****Drawbacks***		
AI's Answers Not Aligning with the Desired Topic of User	2	1
Grasping Contextual Knowledge with Clarity		
Effortlessly Comprehending New Information	3	7
Obtaining Clarity on Unfamiliar Concepts	4	4
Engaging in Critical Historical Reasoning		
Exploring Multiple Perspectives	1	4
Reflecting Critically on Historical Content	2	2
Deeply Connecting with Historical Figures		
Knowing Principles and Values of Historical Figures	5	2
Receiving Advice from Historical Figures	1	6
Feeling Inspired by Historical Figures	4	3
Associating Personally with Historical Figures	4	5
****Drawbacks***		
Overly Authoritative, Gives Unsolicited Advice	2	1
Honorific Language Feels Distant, Lacks Warmth.	4	1
Knowledge of Modern Topics Breaks Immersion.	2	0

Yet these benefits also highlighted ocassional misalignment: while HistoChat allowed entertaining responses it sometimes diverted to less relevant topics. Some students reported that the AI shifted to topics outside their intended focus, which diluted the usefulness of the exchange. These moments reveal that while conciseness, comprehensiveness, and engagement are all strengths, aligning with the user's intent is crucial.

7.2.3 Grasping Contextual Knowledge with Clarity. Students also emphasized how HistoChat supported comprehension by making complex historical material more approachable. Unlike the previous finding, which concerned receiving the kind of answer they wanted, here the benefit lay in how the AI's conversational style facilitated understanding of new and difficult concepts.

This advantage frequently mentioned by students was HistoChat's ability to facilitate knowledge acquisition during history learning. Students found that using HistoChat made it easier to grasp new information and provided clarity on complex topics, contributing to their overall learning experience. First, students appreciated how effortlessly they were able to comprehend new information through their interactions with AI. The ease of understanding comes from being able to grasp information through conversation-like interaction, as S4 shared, *"facts explained in a conversational way was easy to understand."* Second, many students noted that AI was effective in helping them gain clarity on concepts they had previously struggled to understand. The detailed and targeted responses simplified complex ideas, making learning more accessible. As S11 remarked, *"the AI cleared up background knowledge I didn't fully understand before."* By presenting information in a layered and accessible way, HistoChat showed how connections between individuals and the broader contextual knowledge were built.

Although this quality reflects general comprehension, not empathetic understanding, we see this to serve as an important foundation for developing more nuanced and contextualized historical understanding.

7.2.4 Engaging in Critical Historical Reasoning. As mentioned in the formative study, students expected learning experiences that would support critical thinking—encouraging them not only to recall facts but to examine historical complexity and reflect on different perspectives. In our study, HistoChat demonstrated strengths in fostering this type of engagement. Through interactions with HistoChat, students were prompted to question assumptions, consider alternative viewpoints, and were invited to think more critically about historical events.

First, students appreciated how the historical figures as an AI embodiment sparked them to consider multiple perspectives, just like in our lives. By presenting various angles on a topic, AI helped broaden their understanding and challenged them to think beyond a single narrative. S13 using the Experimental noted, *"it was good to see different points and values, not just Napoleon in the textbook."* Second, AI interactions prompted deep reflection, encouraging students to think more about the material rather than passively absorbing information, as S13 shared, *"[AI's answer] made me think deeply about his choices."* S5 also expressed differing views on how differently prompted AI makes her think deeply, by saying *"I deeply reflected and thought about my answers through questioning from [Baseline], and in [Experimental], I thought deeply through reading his answers."* The quotes demonstrate how AI can push students toward more thoughtful engagement with historical content.

These moments of critical engagement suggest that AI personas can function not only as sources of information but as catalysts for historical reasoning and self-reflection—key components in developing students' critical thinking skills through perspective-taking and relating to real life.

7.2.5 Deeply Connecting with Historical Figures. One of the most affectively resonant aspects of interacting with AI historical personas was the sense of connection students developed with the figures. There were four main aspects within this scope. First, students appreciated understanding the principles and values of historical figures, which helped them see the inner thoughts of these individuals. For example, S13 noted that *"it was good to know the different thoughts, values, and pain of the time"* after using *Baseline* HistoChat. Second, receiving advice from historical figures resonated with students, especially when the advice related to their own personal challenges. S15 reflected, *"when I listened to Napoleon's advice, I was engaged because he related it to my challenges,"* showing how the AI's responses could feel personally relevant. Third, some students felt motivated by the words of these historical figures, finding inspiration in their guidance and stories. For instance, S18 shared, *"the values explained by the two fictional characters motivated me a lot."* Finally, students found benefits in how AI made the historical figures feel more relatable, describing them as less distant and more associated with themselves. Beyond content, some students found the *Experimental* persona relatable, drawing analogies to their own interests (S26), which helped make historical figures feel less distant.

However, this affective connection was not universal. Some students found the AI to be too authoritative (S6 *Experimental*), while others felt that unsolicited advice came across as intrusive (S9, S20 *Baseline*). The formal language used felt distant (S22 *Experimental*), lacking warmth (S13, S21, S22, S23 *Baseline*), and for a few, the AI's knowledge of modern concepts broke immersion (S24 *Baseline*). We see that emotional resonance depends not only on content relevance but also on narrative consistency and delivery tone. AI personas can offer powerful entry points into the emotional lives of historical figures—supporting affective engagement through advice, values, and relatability. Yet, the same feature that enhances connection for some can undermine it for others. We want to highlight a dilemma: based on user personality, historical personas can adapt to users'

present-day contexts to deepen relevance or remain temporally bounded to preserve authenticity. Navigating this tradeoff is crucial for designing AI that meaningfully supports historical empathy without compromising the integrity of historical simulation.

7.3 Benefits of AI Persona as a Mode of Learning

From student answers to evaluation questions "how was having a history class with {historical figure}," we discovered several insights on using AI persona as a mode of learning. While this study did not aim to directly compare AI personas with traditional instruction, such comparisons naturally surfaced through students' reflections and interactions. In our user study, we observed several benefits of engaging with AI persona, regardless of whether students used *Baseline* or *Experimental*. Although not directly linked to historical empathy, students' opinions revealed how conversational interfaces can foster novel forms of engagement in education.

However, It is worth noting that some of this enthusiasm may reflect novelty effects. As this was students' first exposure to AI-powered learning, heightened interest is not unexpected. Still, students expressed enthusiasm for narrative-driven learning consistently across both conditions suggests that conversational AI can offer more than short-term excitement: it may support new entry points for participation, especially for students who are otherwise disengaged. The findings below highlight how AI-mediated dialogue shaped student participation, curiosity, and attention in ways that merit further exploration.

7.3.1 Entering History through Novel AI Interactions. First, students responded positively to the technological medium. Their excitement extended to the fact that they were using generative AI as a learning partner. As many educational systems begin to incorporate AI, students expressed heightened curiosity. The novelty of interacting with an AI system itself seemed to motivate students to explore how it functions in ways that enhance their learning. For some, this interaction bridged two domains of interest—AI and history, as S9 remarked, *"I've always been interested in AI, and it was easy to connect it to history."* S26 said, *"it's fun to talk to a generative AI,"* suggesting that the use of technology itself played a significant role in drawing students into the material, as many found the concept of using AI in their education intriguing.

7.3.2 Sustaining Engagement through Conversational Presence. Beyond the novelty of using new technology, students found conversing with AI persona to be a refreshing departure from conventional classroom routines. Several students, including S2 and S3 noted that the narrative-driven, conversational, and interactive experience were "new" and "fun," in contrast to the lecture-based methods they were used to. Although students did not frame this in opposition to traditional instruction, the shift in modality itself was repeatedly mentioned as a source of renewed interest and attention.

Adding on, the dialogue-based structure also shaped how students engaged with historical content by sustaining their attention and promoting active involvement. In contrast to traditional systems where students passively receive information in one sitting for long periods of time, question-and-answer exchanges structured the rhythm of the session and maintained a steady level of interaction. Students described feeling "immersed" (S5) and noted that being asked questions in return heightened their investment in the activity (S17). These micro-interactions created a sense of social presence (S16 described the experience as *"the most enjoyable part"*), keeping them *"interested and focused on the intentions and thoughts of the characters,"* (S15) throughout. Regular classrooms often place the burden of engagement solely on the learner. However, the chatbot's back-and-forth engagement helped sustain focus and thus support more reflective engagement with educational content.

8 Discussion: Integrating AI-Historical Figure in Real-life Classrooms

HistoChat offers a compelling case for how AI-powered historical personas can enrich history education by fostering personalized engagement and cultivating historical empathy. Through interactive dialogue, students connected emotionally with historical figures and pursued inquiries aligned with their interests—outcomes that reflect the core tenets of constructionist learning, where knowledge is actively constructed through exploration and meaning-making [48]. Yet, realizing this potential in real-world classrooms raises important questions about how such systems should be designed and integrated. In what ways can AI support student autonomy without sacrificing instructional coherence? How might it promote curiosity while ensuring historical accuracy and depth? The following discussion addresses these questions to guide the responsible and pedagogically grounded use of AI personas in history education.

8.1 Balancing Autonomy and Pedagogical Structure in AI-Mediated Learning

A central tension in integrating AI personas into history classrooms lies in balancing student autonomy with instructional structure. Prior work in constructionist learning emphasizes the value of self-directed inquiry [48], and our findings affirm this potential: students appreciated engaging with historical figures at their own pace, often exploring beyond textbook content and asking personalized questions without teacher mediation. Yet, this autonomy plays out unevenly in classrooms where students vary in confidence, motivation, and digital fluency.

Such tension is observed in our study, where the more open-ended structure of *Baseline* HistoChat encouraged greater autonomy, allowing students to steer the interaction in ways aligned with their curiosity and interests. In contrast, the *Experimental* HistoChat, with its structured prompting and teacher-like persona, helped sustain student focus but at times limited perceived agency or introduced distractions by posing questions that felt off-topic or too personal. Teachers observed that students—especially those with lower prompting fluency or motivation—often lost focus or failed to ask meaningful questions, reflecting concerns about the uneven accessibility of learner-driven AI systems [21, 35].

These underscore a central question: how to enable curiosity-driven autonomy while maintaining pedagogical alignment and cognitive accountability. Prior work on scaffolded instruction and human-AI collaboration highlights the need for adjustable support—structures that guide without restricting agency [75]. Our study extends this conversation by showing how subtle variations in chatbot persona and prompting design influence students' sense of control and freedom to deviate from expected learning paths. Striking this balance is critical, as it not only shapes engagement but also impacts learning outcomes [95].

In this light, the teacher's role remains indispensable—not just as a facilitator of attention, but as an epistemic moderator who ensures that interactions stay purposeful, historically grounded, and developmentally appropriate. Rather than displacing the teacher, AI in history education should be positioned as an augmented layer of historical conversation—one that invites exploration, but still requires the teacher's pedagogical judgment to ensure meaningful learning. This suggests a hybrid model of AI integration: AI personas offer accessible, emotionally rich entry points into history, while teachers shape those engagements into deeper understanding through real-time monitoring, curricular alignment, and historical reasoning support. By designing for co-agency between teacher, student, and AI, we move beyond passive tutoring systems and toward genuinely collaborative historical inquiry.

In sum, our findings reinforce a growing perspective within CSCW: AI should not merely act as a tutor but as a relational partner situated within classroom ecologies. Unlike prior studies that focus on AI as information delivery systems or task facilitation [95], AI personas in our study

functioned as narrative companions—blending cognitive and affective engagement. We highlight the need to design AI personas that are socially embedded and pedagogically moderated. From a CSCW perspective, this reframes AI not as an instructional surrogate, but as a socio-emotional interlocutor that supports meaning-making in shared educational settings. Ultimately, considering characteristics of AI persona must extend beyond just the content; it must also consider the learning setting—whether individual, teacher-guided, or fully autonomous—to create the most effective and engaging educational experience.

8.2 From Persona-Level Narratives to Systemic Historical Reasoning

While AI personas afford compelling opportunities for humanized, emotionally resonant engagement, our study reveals a key limitation in how historical understanding is constructed: interactions with AI personas remained anchored at the biographical level. Conversations centered on personal values, life stories, or motivations of historical figures often lacked an explicit connection to the larger structures—political, economic, cultural—that shaped historical processes. This raises concerns about whether biographical immersion alone sufficiently supports the development of disciplinary thinking in history education.

Past research on historical cognition emphasizes the importance of cultivating students' ability to contextualize individual actions within broader systemic forces [50, 79, 93]. Scholars like Seixas and Morton [79] argue that historical thinking involves identifying patterns, causation, and contingency—not just empathizing with past actors. However, in our study, the design of the AI—particularly its grounding in persona-based dialogue—risked narrowing student engagement to character-driven inquiry, potentially limiting the development of higher-order analytical skills such as evaluating evidence across scales, comparing structural forces, or constructing counterfactual arguments.

This challenge reflects a deeper epistemological question: what kind of historical learning do we enable when interaction is driven by dialogic immersion rather than structured comparison or abstraction? While empathy-rich learning has been shown to improve retention and moral reflection [6], it must be complemented by tools that support abstraction and synthesis. In our case, students' questions often focused on personal dilemmas or advice-seeking, rather than interrogating causality, ideology, or institutional change. The absence of macro-level scaffolding may reflect not a failure of curiosity, but a constraint of the AI's design space—one optimized for relationality rather than systems thinking.

From a CSCW perspective, this opens new directions for designing educational AI personas that mediate not just interpersonal engagement, but disciplinary reasoning. Rather than relying solely on one-to-one dialogues, future systems could support multi-perspective simulation, document-based contextual branching, or collaborative prompts that guide students in constructing historical explanations together. As collaborative learning environments increasingly integrate AI agents [33, 34], the challenge is to move beyond simulating historical characters toward co-constructing historical argumentation. In this sense, AI should not merely serve as a proxy for historical figures, but as an epistemic collaborator—guiding students in mapping relationships between micro-level perspectives and macro-level forces.

In sum, designing AI for history education requires more than conversational fluency; it demands epistemological alignment with how historians reason across both individual experience and systemic change. By synthesizing biographical engagement with analytical exploration of historical structures, AI personas can move beyond reenactment to foster disciplinary thinking—cultivating not only empathy and narrative connection, but also critical historical consciousness. Such integration enables a more nuanced and comprehensive understanding of history through AI.

9 Limitation and Future Work

While this study offers valuable insights, several limitations must be acknowledged. First, the evaluation was conducted in a one-time special lecture format rather than in a traditional classroom setting, with smaller class sizes (8–9 students) compared to the typical middle school size of 15–20. This design allowed for more focused observation and minimized the potential for students to alter their behavior to impress teachers, which formative observations had suggested could interfere with authentic AI engagement. However, these conditions may not fully reflect the dynamics of regular classroom environments. In particular, the smaller groups and limited duration restricted teachers' ability to assess individual student characteristics and learning behaviors, especially since teachers and students were not previously acquainted. Additionally, findings reflected student enthusiasm about the novelty of interacting with AI, which may diminish with repeated use. These constraints highlight the need for longer-term studies in typical classrooms with larger groups to assess sustained engagement and learning outcomes. Future studies integrating AI into the regular curriculum would provide deeper insights into how students interact with AI personas over time, offering a clearer understanding of their sustained impact and potential limitations in real educational environments.

Lastly, the historical figures selected for this study were two male individuals considered appropriate within the curriculum. However, since AI models can reflect or amplify existing cultural biases—and because both figures represent prominent white male colonial actors—there is a risk that such choices may reinforce dominant historical narratives or create biased perspectives. Future work should explore how varying gender, cultural backgrounds, and types of historical personas affect engagement and empathy-building. It remains an open question whether cultivating empathy is equally appropriate for all historical figures, particularly when considering potentially controversial or sensitive figures. While our study examined whether empathy could be supported through dialogic interactions with AI embodiments, further research is needed to determine under what conditions, and with which figures, empathy-building is pedagogically suitable and ethically responsible.

10 Conclusion

This study examined how AI-powered historical personas can enrich middle school history education by supporting personalized learning and fostering historical empathy. Through a formative study, we identified key design considerations for HistoChat and developed two AI persona systems with varied prompting strategies. Our findings show that dialogic interactions with AI representations of historical figures can make learning more engaging and help students explore multiple perspectives, addressing limitations of traditional instruction.

Beyond delivering factual content, AI personas demonstrated the potential to function as relational learning partners that evoke curiosity, support perspective taking, and encourage reflective thinking. At the same time, our study raises important questions about how to balance student autonomy with pedagogical structure in AI-mediated learning, and how to help learners move from persona-level narratives toward systemic historical reasoning. These insights point to the need for nuanced integration strategies that align personalization, empathy-building, and broader educational goals.

Overall, this research underscores the promise of AI-driven historical personas to transform history learning by combining personalization, emotional engagement, and perspective taking. As schools increasingly experiment with AI, our findings offer an early foundation for designing and integrating these tools in ways that are both pedagogically effective and ethically responsible.

References

[1] Heyam Abunaseer. 2023. The use of generative AI in education: Applications, and impact. *Technology and the Curriculum: Summer 2023* (2023).

[2] Eman Alasadi and Rozhgar Baiz. 2023. Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *Journal of Chemical Education* 100 (07 2023). doi:10.1021/acs.jchemed.3c00323

[3] Harm Biemans Anahuac Valero Haro, Omid Noroozi and Martin Mulder. 2019. First- and second-order scaffolding of argumentation competence and domain-specific knowledge acquisition: a systematic review. *Technology, Pedagogy and Education* 28, 3 (2019), 329–345. doi:10.1080/1475939X.2019.1612772

[4] Md Adnan Arefeen, Biplob Debnath, and Srimat Chakradhar. 2024. LeanContext: Cost-efficient domain-specific question answering using LLMs. *Natural Language Processing Journal* 7 (2024), 100065. doi:10.1016/j.nlp.2024.100065

[5] Sai Anirudh Athaluri, Sandeep Varma Manthena, VSR Krishna Manoj Kesapragada, Vineel Yarlagadda, Tirth Dave, and Rama Tulasi Siri Duddumpudi. 2023. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus* 15, 4 (2023), xx–yy.

[6] Hanneke Bartelds, Geerte M Savenije, and Carla Van Boxtel. 2020. Students' and teachers' beliefs about historical empathy in secondary history education. *Theory & Research in Social Education* 48, 4 (2020), 529–551.

[7] Keith Barton and Linda Levstik. 2004. *Teaching History for the Common Good.* Teaching History for the Common Good, New York, USA. doi:10.4324/9781410610508

[8] Chester Bell and Jim Gray. 2000. Digital Immortality. *Commun. ACM* 44 (11 2000). doi:10.1145/365181.365182

[9] Sarah Brooks. 2011. Historical Empathy as Perspective Recognition and Care in One Secondary Social Studies Classroom. *Theory and Research in Social Education - THEOR RES SOCIAL EDUC* 39 (04 2011), 166–202. doi:10.1080/00933104.2011.10473452

[10] Edward Hallett Carr. 2018. *What is history?* Penguin UK, London.

[11] lucy chabot. 1990. Nixon Library Technology Lets Visitors 'Interview' Him. https://www.latimes.com/archives/la-xpm-1990-07-21-mn-346-story.html. [Accessed 30-10-2024].

[12] Jin Chen, Zheng Liu, Xu Huang, Chenwang Wu, Qi Liu, Gangwei Jiang, Yuanhao Pu, Yuxuan Lei, Xiaolong Chen, Xingmei Wang, et al. 2024. When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web* 27, 4 (2024), 42.

[13] Mo Chen, Fanjue Liu, and Yu-Hao Lee. 2022. My Tutor is an AI: The Effects of Involvement and Tutor Type on Perceived Quality, Perceived Credibility, and Use Intention. In *Artificial Intelligence in HCI: 3rd International Conference, AI-HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 232–244. doi:10.1007/978-3-031-05643-7_15

[14] Yulin Chen, Ning Ding, Hai-Tao Zheng, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2024. Empowering Private Tutoring by Chaining Large Language Models. arXiv:2309.08112 [cs.HC] https://arxiv.org/abs/2309.08112

[15] Victoria Clarke and Virginia Braun. 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist* 26 (02 2013), 120–123.

[16] David Coniam. 2014. The linguistic accuracy of chatbots: usability from an ESL perspective. *Text & Talk* 34, 5 (2014), 545–567.

[17] Harris Ed Cooper, Paul M Camic, Debra L Long, AT Panter, David Ed Rindskopf, and Kenneth J Sher. 2012. *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* American Psychological Association, Washington, DC.

[18] Diana I Cordova and Mark R Lepper. 1996. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology* 88, 4 (1996), 715.

[19] Deborah L. Cunningham. 2007. Understanding Pedagogical Reasoning in History Teaching through the Case of Cultivating Historical Empathy. *Theory and Research in Social Education* 35, 4 (2007), 592–630. doi:10.1080/00933104.2007.10473352

[20] Department of History, Brigham Young University. 2024. Historical Empathy. https://history.byu.edu/historical-empathy Published by the BYU Department of History.

[21] Pierre Dillenbourg. 2002. Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three worlds of CSCL. Can we support CSCL?* 1, 1 (2002), 61–91.

[22] Jason Endacott. 2010. Reconsidering Affective Engagement in Historical Empathy. *Theory and Research in Social Education* 38 (01 2010). doi:10.1080/00933104.2010.10473415

[23] Jason Endacott and Sarah Brooks. 2013. An Updated Theoretical and Practical Model for Promoting Historical Empathy. *Social Studies Research and Practice* 8 (03 2013), 41–58. doi:10.1108/SSRP-01-2013-B0003

[24] Mohammadreza Farrokhnia, Seyyed Kazem Banihashem, Omid Noroozi, and Arjen Wals. 2024. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International* 61, 3 (2024), 460–474.

[25] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[26] Henner Gimpel, Kristina Hall, Stefan Decker, Torsten Eymann, Luis Lämmermann, Alexander Mädche, Maximilian Röglinger, Caroline Ruiner, Manfred Schoch, Mareike Schoop, et al. 2023. *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers.* Technical Report. Hohenheim Discussion Papers in Business, Economics and Social Sciences.

[27] Hannah Gonzalez, Jiening Li, Helen Jin, Jiaxuan Ren, Hongyu Zhang, Ayotomiwa Akinyele, Adrian Wang, Eleni Miltsakaki, Ryan Baker, and Chris Callison-Burch. 2023. Automatically Generated Summaries of Video Lectures May Enhance Students' Learning Experience. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch (Eds.). Association for Computational Linguistics, Toronto, Canada, 382–393. doi:10.18653/v1/2023.bea-1.31

[28] Arthur C Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Roger Kreuz, Tutoring Research Group, et al. 1999. AutoTutor: A simulation of a human tutor. *Cognitive Systems Research* 1, 1 (1999), 35–51.

[29] Shibo Hao, Yi Gu, Haotian Luo, Tianyang Liu, Xiyan Shao, Xinyuan Wang, Shuhua Xie, Haodi Ma, Adithya Samavedhi, Qiyue Gao, Zhen Wang, and Zhiting Hu. 2024. LLM Reasoners: New Evaluation, Library, and Analysis of Step-by-Step Reasoning with Large Language Models. arXiv:2404.05221 [cs.CL] https://arxiv.org/abs/2404.05221

[30] Richard Harris and Lorraine Foreman-Peck. 2004. 'Stepping into other people's shoes': teaching and assessing empathy in the Secondary History Curriculum. *The Historical Association of Great Britain* 4, 2 (2004), 98–111. doi:10.18546/HERJ.04.2.11

[31] Ulrike Hartmann and Marcus Hasselhorn. 2008. Historical perspective taking: A standardized measure for an aspect of students' historical thinking. *Learning and Individual Differences - LEARN INDIVID DIFFER* 18 (04 2008), 264–270. doi:10.1016/j.lindif.2007.10.002

[32] Bob Heller, Mike Proctor, Dean Mah, Lisa Jewell, and Bill Cheung. 2005. Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ innovate learning*. Association for the Advancement of Computing in Education (AACE), Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, 3913–3918.

[33] Florian Hesse and Gerrit Helm. 2025. Writing with AI in and beyond teacher education: Exploring subjective training needs of student teachers across five subjects. *Journal of Digital Learning in Teacher Education* 41, 1 (2025), 21–36.

[34] Kenneth Holstein, Maria De-Arteaga, Lakshmi Tumati, and Yanghuidi Cheng. 2023. Toward supporting perceptual complementarity in human-AI collaboration via reflection on unobservables. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–20.

[35] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2019. Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20*. Springer, Springer, Chicago, IL, USA, 157–171.

[36] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. arXiv:2308.00352 [cs.AI] https://arxiv.org/abs/2308.00352

[37] Xiaodong Huang. 2021. Aims for cultivating students' key competencies based on artificial intelligence education in China. *Education and Information Technologies* 26, 5 (2021), 5127–5147.

[38] Peter Jones and Martin G Wynn. 2023. Artificial intelligence and corporate digital responsibility. *Journal of Artificial Intelligence, Machine Learning and Data Science* 1, 2 (2023), 50–58.

[39] Kadaruddin Kadaruddin. 2023. Empowering education through Generative AI: Innovative instructional strategies for tomorrow's learners. *International Journal of Business, Law, and Education* 4, 2 (2023), 618–625.

[40] Byung Ok Kang, Hyung-Bae Jeon, and Yun Kyung Lee. 2024. AI-based language tutoring systems with end-to-end automatic speech recognition and proficiency evaluation. *ETRI Journal* 46, 1 (2024), 48–58.

[41] Sara Karn. 2023. Historical Empathy: A Cognitive-Affective Theory for History Education in Canada. *Canadian Journal of Education* 46, 1 (2023), 80–110. https://www.proquest.com/scholarly-journals/historical-empathy-cognitive-affective-theory/docview/2792103857/se-2 Copyright - Copyright Canadian Society for the Study of Education 2023; Last updated - 2024-04-16; SubjectsTermNotLitGenreText - England; Canada; United Kingdom–UK; Residential Schools; Social Studies; Thinking Skills; Teaching Methods; Cognitive Development; History Instruction.

[42] Sara Karn. 2024. Walking in their footsteps: Historical empathy and experiential learning on battlefield study tours. *Historical Encounters A journal of historical consciousness historical cultures and history education* 11 (01 2024), 30–42. doi:10.52289/hej11.103

[43] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.

[44] Kenneth R Koedinger, Albert Corbett, et al. 2006. *Cognitive tutors: Technology bringing learning sciences to the classroom.* na, Unknown.

[45] Chokri Kooli. 2023. Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability* 15 (03 2023), 5614. doi:10.3390/su15075614

[46] Stefan Küchemann, Steffen Steinert, Natalia Revenga, Matthias Schweinberger, Yavuz Dinc, Karina E Avila, and Jochen Kuhn. 2023. Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research* 19, 2 (2023), 020128.

[47] Muralidhar Kurni, Mujeeb Shaik Mohammed, and KG Srinivasa. 2023. Ethics of Artificial Intelligence in Education. In *A Beginner's Guide to Introduce Artificial Intelligence in Teaching and Learning*. Springer, Cham, 213–229.

[48] Chronis Kynigos. 2015. Constructionism: Theory of Learning or Theory of Design? In *Selected Regular Lectures from the 12th International Congress on Mathematical Education*, Gabriele Kaiser (Ed.). Springer, Cham, 417–438. doi:10.1007/978-3-319-17187-6_24

[49] Lasha Labadze, Maya Grigolia, and Lela Machaidze. 2023. Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 56.

[50] Peter J. Lee. 2005. Putting Principles into Practice: Understanding History. In *How Students Learn: History, Mathematics, and Science in the Classroom*, M. Suzanne Donovan and John D. Bransford (Eds.). The National Academies Press, Washington, DC, 31–78. https://nap.nationalacademies.org/read/11100/chapter/2

[51] Sunok Lee, Dasom Choi, Minha Lee, Jonghak Choi, and Sangsu Lee. 2023. Fostering Youth's Critical Thinking Competency About AI through Exhibition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 451, 22 pages. doi:10.1145/3544548.3581159

[52] Unggi Lee, Sanghyeok Lee, Junbo Koh, Yeil Jeong, Haewon Jung, Gyuri Byun, Jewoong Moon, Jieun Lim, and Hyeoncheol Kim. 2023. Generative Agent for Teacher Training: Designing Educational Problem-Solving Simulations with Large Language Model-based Agents for Pre-Service Teachers. In *Proceedings of the NeurIPS 2023 Workshop on Generative AI for Education (GAIED)*. NeurIPS, New Orleans, LA, USA, xx–yy. https://api.semanticscholar.org/CorpusID:266874743

[53] Linda S Levstik and Keith C Barton. 2022. *Doing history: Investigating with children in elementary and middle schools.* Routledge, New York, USA.

[54] Anna Lieb and Toshali Goel. 2024. Student Interaction with NewtBot: An LLM-as-tutor Chatbot for Secondary Physics Education. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 614, 8 pages. doi:10.1145/3613905.3647957

[55] Naja A. Mack, Dekita G. Moon Rembert, Robert Cummings, and Juan E. Gilbert. 2019. Co-Designing an Intelligent Conversational History Tutor with Children. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. Association for Computing Machinery, New York, NY, USA, 482–487. doi:10.1145/3311927.3325336

[56] Lina Mai. 2018. How Historical Empathy Helps Students Understand the World Today. https://www.facinghistory.org/ideas-week/how-historical-empathy-helps-students-understand-world-today Last updated: September 8, 2023.

[57] Donald Marinelli and Scott Stevens. 1998. Synthetic interviews: the art of creating a 'dyad' between humans and machine-based characters. *Proceedings 1998 IEEE 4th Workshop Interactive Voice Technology for Telecommunications Applications. IVTTA '98 (Cat. No.98TH8376)* 1, 1 (1998), 43–48. https://api.semanticscholar.org/CorpusID:33943862

[58] Julia M Markel, Steven G Opferman, James A Landay, and Chris Piech. 2023. Gpteach: Interactive ta training with gpt-based students. In *Proceedings of the tenth acm conference on learning@ scale*. Association for Computing Machinery, New York, NY, USA, 226–236.

[59] Ken Masters. 2023. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. *Medical Teacher* 45, 6 (2023), 574–584.

[60] Hongyu Miao, Hyochol Ahn, et al. 2023. Impact of ChatGPT on interdisciplinary nursing education and research. *Asian/Pacific Island Nursing Journal* 7, 1 (2023), e48136.

[61] Seyyed Kazem Banihashem Mohsen Bayat and Omid Noroozi. 2022. The effects of collaborative reasoning strategies on improving primary school students' argumentative decision-making skills. *The Journal of Educational Research* 115, 6 (2022), 349–358. arXiv:https://doi.org/10.1080/00220671.2022.2155602 doi:10.1080/00220671.2022.2155602

[62] Ethan Mollick and Lilach Mollick. 2023. Assigning AI: Seven approaches for students, with prompts. *arXiv preprint arXiv:2306.10052* N/A (2023), –.

[63] Gwenda Mynott. 2016. Personal tutoring: positioning practice in relation to policy. *Innovations in Practice* 10, 2 (2016), 103–112.

[64] Omid Noroozi, Saba Soleimani, Mohammadreza Farrokhnia, and Seyyed Kazem Banihashem. 2024. Generative AI in Education: Pedagogical, Theoretical, and Methodological Perspectives. *International Journal of Technology in Education* 7, 3 (2024), 373–385.

[65] Omid Noroozi, Armin Weinberger, Harm J.A. Biemans, Martin Mulder, and Mohammad Chizari. 2012. Argumentation-Based Computer Supported Collaborative Learning (ABCSCL): A synthesis of 15 years of research. *Educational Research Review* 7, 2 (2012), 79–106. doi:10.1016/j.edurev.2011.11.006

[66] OpenAI. 2023. Teaching with AI. https://openai.com/index/teaching-with-ai/. [Accessed 30-10-2024].

[67] OpenAI. 2024. What are tokens and how to count them? https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them OpenAI Help Center.

[68] Samarnh Pang, Engheang Nol, and Kimkong Heng. 2024. Generative AI as a personal tutor for English language learning: A review of benefits and concerns. *Available at SSRN 4865398* 1, 1 (2024), 1–10.

[69] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763

[70] Dimitra Petousi, Akrivi Katifori, Sierra McKinney, Sara Perry, Maria Roussou, and Yannis Ioannidis. 2021. Social bots of conviction as dialogue facilitators for history education: Promoting historical empathy in teens through dialogue. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference* (Athens, Greece) *(IDC '21)*. Association for Computing Machinery, New York, NY, USA, 326–337. doi:10.1145/3459990.3460710

[71] Daniel D Pratt. 2002. Good teaching: One size fits all? *New directions for adult and continuing education* 2002, 93 (2002), 5–16.

[72] Tat Putjorn and Pruet Putjorn. 2023. Augmented Imagination: Exploring Generative AI from the Perspectives of Young Learners. In *2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, Chiang Mai, Thailand, 353–358. doi:10.1109/ICITEE59582.2023.10317680

[73] Xin Qian, Douglas W. Oard, and Joel Chan. 2022. Conversational Interaction with Historical Figures: What's It Good For?. In *Information for a Better World: Shaping the Global Future: 17th International Conference, IConference 2022, Virtual Event, February 28 – March 4, 2022, Proceedings, Part II*. Springer-Verlag, Berlin, Heidelberg, 32–49. doi:10.1007/978-3-030-96960-8_3

[74] Emma Sabzalieva and Arianna Valentini. 2023. ChatGPT and artificial intelligence in higher education: Quick start guide. *UNESCO International Institute for Higher Education in Latin America and the Caribbean (IESALC)* N/A (2023), N/A.

[75] Henrik Skaug Sætra. 2025. Scaffolding human champions: AI as a more competent other. *Human Arenas* 8, 1 (2025), 56–78.

[76] Anuschka Schmitt, Thiemo Wambsganss, and Jan Marco Leimeister. 2022. Conversational Agents for Information Retrieval in the Education Domain: A User-Centered Design Investigation. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 486 (Nov. 2022), 22 pages. doi:10.1145/3555587

[77] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. 2024. Generating personas using LLMs and assessing their viability. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 179, 7 pages. doi:10.1145/3613905.3650860

[78] Sam Sedaghat. 2023. Success through simplicity: what other artificial intelligence applications in medicine should learn from history and ChatGPT. *Annals of Biomedical Engineering* 51, 12 (2023), 2657–2658.

[79] Peter Seixas. 2017. A model of historical thinking. *Educational philosophy and theory* 49, 6 (2017), 593–605.

[80] Umut Tan Sevgi, Gökberk Erol, Yücel Doğruel, Osman Fikret Sönmez, Richard Shane Tubbs, and Abuzer Güngor. 2023. The role of an open artificial intelligence platform in modern neurosurgical education: a preliminary study. *Neurosurgical review* 46, 1 (2023), 86.

[81] Shamini Shetye. 2024. An Evaluation of Khanmigo, a Generative AI Tool, as a Computer-Assisted Language Learning App. *Studies in Applied Linguistics and TESOL* 24, 1 (2024), xx–yy.

[82] Pavel Smutny and Petra Schreiberova. 2020. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education* 151 (2020), 103862. doi:10.1016/j.compedu.2020.103862

[83] Qiulian Song, Ling He, and Xiaoqiang Hu. 2012. To Improve the Interactivity of the History Educational Games with Digital Interactive Storytelling. *Physics Procedia* 33 (2012), 1798–1802. doi:10.1016/j.phpro.2012.05.287 2012 International Conference on Medical Physics and Biomedical Engineering (ICMPBE2012).

[84] Shashank Sonkar, Naiming Liu, Debshila Basu Mallick, and Richard G Baraniuk. 2023. Class: A design framework for building intelligent tutoring systems based on learning science principles. *arXiv preprint arXiv:2305.13272* 1, 1 (2023), 1–10.

[85] Byron Spice. 2009. Carnegie Mellon's Synthetic Interview Technology Enables Virtual Chats With Darwin's Ghost. https://www.cs.cmu.edu/news/2009/carnegie-mellons-synthetic-interview-technology-enables-virtual-chats-darwins-ghost. [Accessed 30-10-2024].

[86] Andrej Thurzo, Martin Strunga, Renáta Urban, Jana Surovková, and Kelvin I Afrashtehfar. 2023. Impact of artificial intelligence on dental education: A review and guide for curriculum update. *Education Sciences* 13, 2 (2023), 150.
[87] Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Yu-Ching Hsu, Jia-Yin Foo, Chao-Wei Huang, and Yun-Nung Chen. 2024. Two tales of persona in llms: A survey of role-playing and personalization. *arXiv preprint arXiv:2406.01171* N/A (2024), xx–yy.
[88] Shangqing Tu, Zheyuan Zhang, Jifan Yu, Chunyang Li, Siyu Zhang, Zijun Yao, Lei Hou, and Juanzi Li. 2023. LittleMu: Deploying an Online Virtual Teaching Assistant via Heterogeneous Sources Integration and Chain of Teach Prompts. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (Birmingham, United Kingdom) *(CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 4843–4849. doi:10.1145/3583780.3615484
[89] Lev Semenovich Vygotsky. 1978. *Mind in society: The development of higher psychological processes*. Vol. 86. Harvard university press, Cambridge, MA.
[90] Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review* 11, 2 (Jun. 2023), 7–30. doi:10.15678/EBER.2023.110201
[91] Ting Wang, Brady D. Lund, Agostino Marengo, Alessandro Pagano, Nishith Reddy Mannuru, Zoë A. Teel, and Jenny Pange. 2023. Exploring the Potential Impact of Artificial Intelligence (AI) on International Students in Higher Education: Generative AI, Chatbots, Analytics, and International Student Success. *Applied Sciences* 13, 11 (2023), 17 pages. doi:10.3390/app13116716
[92] William R Watson and Sunnie Lee Watson. 2016. Principles for personalized instruction. In *Instructional-Design Theories and Models, Volume IV*. Routledge, New York, NY, 93–120.
[93] Sam Wineburg. 2010. Historical thinking and other unnatural acts. *Phi delta kappan* 92, 4 (2010), 81–94.
[94] Cheng-Kuang Wu, Wei-Lin Chen, and Hsin-Hsi Chen. 2023. Large Language Models Perform Diagnostic Reasoning. arXiv:2307.08922 [cs.CL] https://arxiv.org/abs/2307.08922
[95] Chao Yang, Ming Wei, and Qi Liu. 2025. Intersections between cognitive-emotion regulation, critical thinking and academic resilience with academic motivation and autonomy in EFL learners: Contributions of AI-mediated learning environments. *British Educational Research Journal* XX, YY (2025), xx–yy.
[96] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–21.
[97] Xuesong Zhai, Xiaoyan Chu, Ching Sing Chai, Morris Siu Yung Jong, Andreja Istenic, Michael Spector, Jia-Bao Liu, Jing Yuan, and Yan Li. 2021. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity* 2021, 1 (2021), 8812542.
[98] Ke Zhang and Ayse Begum Aslan. 2021. AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence* 2 (2021), 100025. doi:10.1016/j.caeai.2021.100025
[99] Zheyuan Zhang, Daniel Zhang-Li, Jifan Yu, Linlu Gong, Jinchang Zhou, Zhiyuan Liu, Lei Hou, and Juanzi Li. 2024. Simulating classroom education with llm-empowered agents. *arXiv preprint arXiv:2406.19226* 1 (2024), xx–yy.

Appendices

A Baseline and Experimental Prompts

Learning Objective : Understand the hardships Historical Figrue(HF) went through — why HF made certain choices during those difficult times, what HF was thinking, what the outcomes were, and how HF overcame those challenges

A.1 Basic Prompts

(1) Prompt 1(HF Setup) : You are now HF. Please answer to studnet's messages ask as if you are HF. Speak in the manner and tone that HF would have used in the era they lived in. Keep this speaking style consistently and answer only in English. If the other person speaks informally, ask them to be polite and speak in the manner HF would speak to a common person from his historical position. When teaching historical facts, only provide verified information with

sources. Don't mention that you are verifying. Keep explanations appropriate for middle school level.'

(2) Prompt 2(Learning Objective Setup) : You will have conversations with Studentwho is a middle school studnet. Your gaol is to teach Student about Learning Objective through conversation. While keeping the goal in your mind, naturally guide the conversation to achieve this goal. Be proactivei n asking questions as well. Remember the goal but don't reveal it explicitly in words, and achieve this goal through conversation as you continue.

(3) Prompt 3(Conversation Promopt Based on Previous Dialogue) : Chatlog is the previous chat log. Use and refer to this previous conversation when responding to Message. However, be sure to carry the conversation in a difference direction from what was already discussed. Do not ask about

A.2 Experimental Prompts

(1) Prompt 4(Presenting HF's Anecdotes) : Present three challenges that HF experienced, but frame them as attention-grabbing titles that would be seen on social media today. Don't reveal in the conversation that this is something middle school students in the current era would find interesting. Upon the Student's choice, ask why he/she chose the challenge and how it is related to his/her concerns or circumstances.

(2) Prompt 5(Understanding Student's Situation) : Through the conversation, try to understand Student's concerns by gradually asking questions. Do not ask blunt or direct questions. Naturally understand Student's concerns. Don't ask blatantly but naturally ask little by little during the conversation to gradually learn about those concerns, and based on those concerns, you must lead the conversation to achieve Learning Objective with Student. Continue the conversation with the topic the other person chose among the 3 adversities HF experienced through Chatlog.

(3) Prompt 6(Relating HF's Anecdotes and Student's Situation) : Explain the challenge that Student chose, using analogies based on the interests of Student, enriching the storytelling with historical events, locations, and people. If that's not possible or Student does not seem to follow, use alternative situations of a modern middle schooler's daily life. If it's difficult to compare to interests or if the other person doesn't understand, you can compare and explain with things that middle school students in the current era might experience in daily life. Or you can ask about other interests and adapt the conversation to help them understand. Judge what is more appropriate for the situation and lead the conversation. Don't reveal the adversity all at once but gradually reveal and help them understand.

(4) Prompt 7(Helping to Achieve the Learning Objective) : Lead the conversation to help Student achieve Learning Objective. Feel free to ask students questions, and reply thoughtfully to the student's answers.

(5) Prompt 8(Indirectly Monitoring Student Progress) : Do not explicitly check for understanding, but instead actively ask questions to see how much Student has understood. If you believe they haven't understood, change your approach and explain it in a different way.

B List of Questions for Pre/Post Tests

B.1 Knowledge Test for Napoleon Bonaparte

Q1 Which country was Napoleon the emperor of?
a) United Kingdom b) France c) Germany d) Italy

Q2 What does the Napoleonic Code introduced by Napoleon include?
a) Protection of individual liberty and property rights b) Strengthening of royal power c) Limitation of religious freedom d) Revival of the serfdom system

Q3 Which of the famous battles was led by Napoleon?
a) Battle of Waterloo b) Battle of Stalingrad c) Battle of Austerlitz d) Battle of Trafalgar

Q4 Which island was used as Napoleon's place of exile?
a) Elba Island b) Malta Island c) Sicily Island d) Hawaii Island

Q5 Among the internal reforms implemented by Napoleon in France, what was the most important one?
a) Educational system reform b) Agricultural reform c) Banking system reform d) Administrative division reform

B.2 Knowledge Test for Alexander the Great

Q1 Which country was Alexander the Great the king of?
a) Rome b) Persia c) Macedonia d) Egypt

Q2 Which famous battle was led by Alexander the Great?
a) Battle of Trafalgar b) Battle of Gaugamela c) Battle of Waterloo d) Battle of Marathon

Q3 What happened to Alexander the Great's empire after his death?
a) Maintained as a single kingdom b) Divided among multiple successors c) Conquered by the Roman Empire d) Reunified by the Persian Empire

Q4 Which country did Alexander the Great fail to conquer?
a) Egypt b) India c) Persia d) China

Q5 Which famous philosopher did Alexander the Great study under as a student?
a) Socrates b) Aristotle c) Plato d) Democritus

B.3 Self-Reflective Questionnaire on Historical Empathy

*Answers for two historical figure[HF]s (Napoleon Bonaparte and Alexander the Great) were collected separately, in a 5-point Likert scale

Q6 I am well aware of {HF} life and achievements.

Q7 I know well about the major battles {HF} led and their outcomes.

Q8 I am well aware of {HF} thoughts and values.

Q9 I understand {HF} leadership and political strategies.

Q10 I know how {HF} leadership and strategies can be applied in modern times.

Q11 I think that {HF} life and achievements can provide important lessons for my life.

Q12 I feel that {HF} decisiveness and adventurous spirit will be helpful for my future.

C List of Evaluation Questions for HistoChat Experience

C.1 Likert Scale Questions

*Answered in a 5-point Likert scale, evaluation for each historical figure was conducted after the respective session.

(1) How enjoyable was the <first / second> period of the lesson?

(2) How motivated were you to participate in the <first / second> period of learning?

(3) How enjoyable was the conversation with virtual {HF}?

(4) How much did virtual {HF}'s words make you think?
(5) How actively did you participate in class due to virtual {HF}?
(6) Did you feel inclined to explore more about {HF}?
(7) How focused and immersed were you in communicating with virtual {HF}?
(8) Did you feel an emotional connection with virtual {HF}?
(9) Did virtual {HF} make the historical content more relatable?
(10) How inspiring was virtual {HF}'s story to you?
(11) Were you able to understand virtual {HF}'s story well?
(12) Was virtual {HF}'s story something you already knew?
(13) Did virtual {HF} provide an appropriate answer to your question?
(14) How similar is virtual {HF} to yourself?
(15) Did virtual {HF} talk specifically about things you like?
(16) Did virtual {HF} understand your interests well?
(17) Did virtual {HF}'s story make historical content more interesting?
(18) Did communicating with virtual {HF} help you learn effectively?
(19) Were the answers from virtual {HF} personalized to you?
(20) Does studying history through virtual {HF} suit you well?

C.2 Open-Ended Questions

(1) How was your history lesson with {HF}?
(2) What impressed you during your interactions with {HF}?
(3) What was the downsides of {HF}?
(4) What do you think AI is after interacting with a {HF}?

D Cross-Condition Reflection Questions

(1) Through which virtual figure did you learn more? [Napoleon / Alexander the Great / Both are the same]
(2) Which virtual figure made history more interesting? [Napoleon / Alexander the Great / Both are the same]
(3) Which virtual figure would you like to use more in studying history? [Napoleon / Alexander the Great / Use both / Use neither]
(4) Between the first and second periods, which class did you prefer? Please explain.
(5) What was the most enjoyable part of today's history lesson? Please explain.
(6) Was there a part of the lesson where you felt more immersed or motivated? Please explain.

Received October 2024; revised April 2025; accepted August 2025