Maike: Designing a Socratic Educational Chatbot to Foster Critical Thinking and Learner Agency

Lucile Favero
lucile@ellisalicante.org
ELLIS Alicante
Alicante, Spain

Juan Antonio Pérez-Ortiz
Universitat d'Alacant
San Vicente del Raspeig, Spain

Tanja Käser
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

Nuria Oliver
ELLIS Alicante
Alicante, Spain

Abstract

Generative AI tools are increasingly used by students for writing support, yet many existing systems risk promoting cognitive offloading, dependency, and homogenized thinking by providing direct answers or polished content. In this paper, we present Maike, a student-centric educational chatbot designed to foster critical thinking through Socratic dialogue rather than content generation.

Maike combines four modules: (1) argument mining, (2) pedagogical strategy selection, (3) critical question generation, and (4) Socratic dialogue.

Instead of polished, pre-packaged text, Maike poses targeted, thought-provoking questions to encourage students to articulate assumptions and evidence, and refine arguments.

We describe Maike's architecture, whose main technical contribution is the integration of argument mining, pedagogical strategy selection, critical question generation, and Socratic dialogue, and discuss its design principles across four dimensions for responsible educational AI: cognition, agency, emotion, and ethics. Our approach emphasizes desirable difficulties, learner autonomy, and privacy-preserving deployment via small open-source language models running locally. In future work, we plan to evaluate Maike in classroom studies to assess learning outcomes and its impact on the students' independent writing and critical thinking skills. This work describes a concrete example of how generative AI can be designed as a cognitive scaffold rather than a cognitive substitute, offering practical guidance for human-centered AI tools in education.

ACM Reference Format:
Lucile Favero, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2026. Maike: Designing a Socratic Educational Chatbot to Foster Critical Thinking and Learner Agency. In *Proceedings of (CHI'26 Tools For Thought).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

CHI'26 Tools For Thought, Barcelona, Spain

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Generative AI is rapidly entering educational contexts, offering new forms of writing support but also leading to cognitive offloading, dependency, and homogenized thinking [7].

Most existing systems prioritize fluency and answer quality, positioning students as passive recipients rather than active thinkers [10]. In the context of Tools for Thought, a central question emerges: how can generative AI support learning without replacing human thought?

We argue that educational AI should act as a cognitive scaffold rather than a cognitive substitute; preserving productive struggle, learner agency, and reflective reasoning. In this paper, we present Maike, a student-centered educational chatbot that supports essay writing through Socratic dialogue. Instead of generating content, Maike poses targeted critical questions grounded in automatic argument analysis and pedagogical strategies, helping students articulate assumptions, examine evidence, and iteratively strengthen their arguments.

Maike integrates four modules (argument mining, pedagogical strategy selection, critical question generation, and Socratic dialogue) to support the thinking process rather than deliver content.

We contribute with a concrete design instantiation of generative AI used as a Tool for Thought, articulated through five principles: (1) question-first interaction, (2) intentional pedagogical friction, (3) dialogue over delivery, (4) argument-aware prompting, and (5) local-first deployment. We conclude by outlining future classroom studies of Maike to assess learning outcomes and its impact on the students' independent writing and critical thinking.

2 Maike as a tool for thought

Maike [5] is an educational chatbot that supports students in developing their essay writing skills through Socratic dialogue [11]. Relative to more conventional ITS formulations surveyed in Zerkouk et al. [15], Maike adopts a dialogic, question-first formulation, while adding argument-aware analysis to steer the interaction. Its goal is twofold: first, to assist students in improving the quality of their essays, and second to cultivate a habit of engaging with topics in a critical, structured, and cogent manner.

Design principles for tools for thought. We propose five principles—referred to with the acronym QALDI—for designing generative AI tools as cognitive scaffolds rather than cognitive substitutes, explicitly positioning Maike as a Tool for Thought:

(1) **Question-first interaction** (Q), prioritizing asking targeted questions over providing answers;
(2) **Argument-aware interaction** (A), grounding prompts in the learners' own reasoning to support metacognition and revision;
(3) **Local-first architectures** (L), which reinforce learner agency, privacy, and equitable access;
(4) **Dialogue over delivery** (D), framing AI as a conversational partner that supports sense-making rather than a source of polished content; and
(5) **Intentional pedagogical friction** (I), introducing desirable difficulties that promote reflection and durable learning;

The QALDI principles place AI tools as a medium for thinking, reflective reasoning, learner autonomy, and intellectual ownership rather than optimizing for convenience, engagement or fluency. They are intended as design guides rather than objectives to maximize simultaneously, and must therefore be balanced in context, for instance between local deployment, analysis quality, and the amount of pedagogical friction appropriate for a given learner.

`Maike`*'s architecture.* We adopt the QALDI principles in `Maike`, whose architecture is depicted in Fig 1.

1. Argument Mining: This module analyzes the student's essay and automatically segments it into arguments, classifying the argument components into three types (claims, premises and counterarguments) [6, 8]. It also evaluates the quality of each argument component, and the overall quality of the essay, based on three complementary dimensions: logical cogency, reasonableness, and rhetorical effectiveness [13]. It is implemented by means of open-source, small LLMs which achieve state-of-the-art performance while preserving the students' privacy and enabling ease of access.

2. Pedagogical Strategy: Based on the output of the argument mining module, this component determines an appropriate pedagogical strategy. Concretely, it maps detected weaknesses to tutoring moves such as asking for clarification, requesting supporting evidence, eliciting counterarguments, or probing underlying assumptions. The choice is conditioned by the type and quality of the detected argument component, and is passed to the downstream question generation module to control both the critical question posed and the subsequent dialogue focus.

3. Critical Question Generation: Using both the argument analysis and the selected pedagogical strategy, this module generates a set of critical questions designed to challenge the student's reasoning. It then selects the most pedagogically relevant question to present [3]. This module is also implemented by means of open-source, small LLMs.

4. Socratic Dialogue: Engaging the student in a Socratic dialogue [11] centered around the selected critical question by means of an open-source small LLM, `Maike` dynamically adapts the conversation based on the student's responses, continuing the exchange until a satisfactory answer is provided. The improved reasoning can then be integrated back into the student's essay [4].

3 Responsible AI in Education

The previously described QALDI principles shape `Maike`'s design and motivate a responsible use of AI in education to preserve thinking, learner ownership, and reflective engagement while reducing risks related to cognitive offloading, dependency, privacy, and academic integrity. In this section, we describe these design intentions across four dimensions (cognitive, agentic, emotional, and ethical), summarized in Table 1. The cognitive dimension is the most directly instantiated in the current architecture, whereas the broader claims regarding agency, emotion, and ethics should be read as hypotheses to be examined in future classroom studies.

3.1 Cognition

Regarding the cognition dimension, `Maike` implements three pedagogical design principles to support cognitive development instead of replacing it.

Fostering Critical Thinking Rather Than Enabling Cognitive Offloading. Unlike many AI tools that provide pre-digested answers, `Maike` is designed to stimulate cognitive effort by posing personalized and targeted critical questions. This approach actively counters the tendency for cognitive offloading and encourages intellectual resilience. Instead of automating thought, `Maike` prompts users to engage in reflective analysis, thus fostering sustained critical thinking. For instance, instead of simply responding with an improved version of the essay, `Maike` might ask, *"What counterarguments might challenge this position?"* or *"Why do you think this evidence is convincing?"* These kinds of prompts require the student to pause, consider multiple perspectives, and make their reasoning explicit, key components of critical thinking, counteracting the overreliance on automated answers and promoting intellectual resilience. By avoiding shortcuts and encouraging students to do the cognitive work themselves, `Maike` supports the development of analytical habits that extend beyond a single interaction.

Pedagogically Grounded Design. `Maike` is grounded in established educational theories and promotes active learning by prompting students to think, reflect, and respond rather than passively receive information. Instead of stating that an argument is weak, it asks questions such as *"What evidence could strengthen this claim?"*, initiating Socratic dialogue that supports revision and reasoning. Drawing on constructivist principles, `Maike` builds on students' existing ideas: rather than overriding partially formed thoughts, it poses follow-up questions like *"How does this connect to your earlier point?"* or *"Can you clarify the relationship between these two ideas?"* to help refine thinking. It also provides adaptive scaffolding by tailoring questions to the student's inferred level of understanding, posing nuanced challenges (*e.g., "What assumptions underlie your argument?"*) for advanced students and simpler prompts (*e.g., "What is your main point here?"*) for beginners; supporting progressive development toward independent thinking.

Desirable Difficulties. `Maike` is designed to counteract the effort-avoidance tendency reinforced by many AI tools. Rather than prioritizing fluency or user satisfaction, it introduces pedagogical friction (intentional cognitive effort) as a central feature of the learning process. Through personalized, Socratic questioning (*e.g., "What assumption does this rely on?"*, *"How does this counterpoint challenge your claim?"*), `Maike` prompts students to retrieve, reassess, and refine their thinking. These questions are calibrated to stretch the student's reasoning without overwhelming them, based on the content of their essay. This approach reframes mental effort not

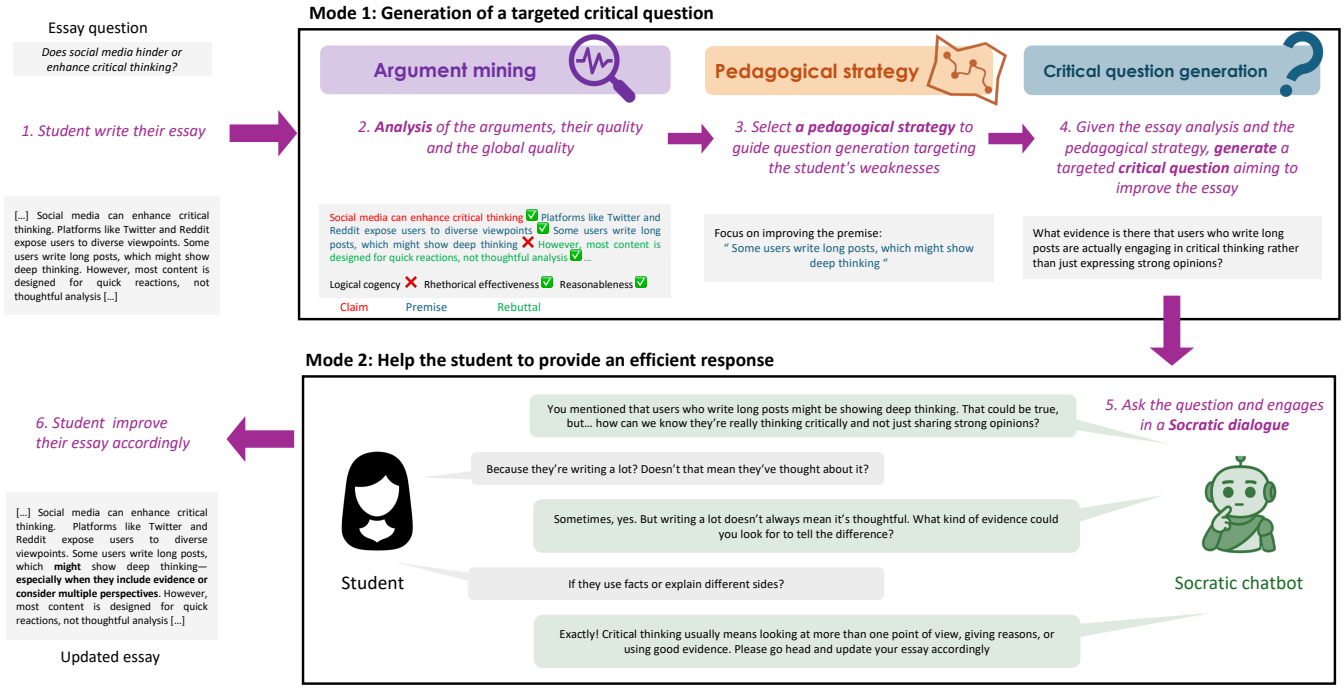

Figure 1: Maike's architecture with its four modules: argument mining (purple), pedagogical strategy selection (orange), critical question generation (blue) and Socratic dialogue (green). Given an essay written by the student (step 1), Maike extracts its arguments, their type and quality (step 2); selects a pedagogical strategy to guide question generation (step 3); generates a targeted critical question (step 4); asks the question and engages in a Socratic dialogue (step 5) to help the student improve their essay (step 6) in an iterative manner.

as failure, but as a signal of meaningful engagement. By sustaining productive difficulty, Maike fosters long-term retention, critical thinking, and resistance to cognitive offloading; avoiding simplification in favor of intellectual growth.

Table 1: Maike's pedagogical design principles regarding the cognitive, agency, emotional, and ethical dimensions

Dimension	Key Idea
Cognition	
Foster critical thinking	Counters cognitive offloading by prompting reflection and analysis; encourages resilience and independent reasoning.
Pedagogically grounded design	Based on constructivist principles and scaffolding; adapts to student's level; promotes active learning through Socratic questioning.
Desirable difficulties	Introduces productive cognitive effort; reframes struggle as meaningful engagement; fosters retention and intellectual growth.
Agency	
Autonomous thinking	Aims to mitigate overreliance on AI; positions chatbot as a partner in inquiry; supports autonomy and metacognitive awareness
Mitigation of knowledge homogenization	Aims to encourage diverse perspectives and original thinking; reduces convergence towards standardized AI-shaped responses.
Emotion	
Self-efficacy regulation	Socratic dialogue is intended to support self-efficacy, confidence, and ownership of reasoning.
Mitigates inadequacy	May reduce AI-related discomfort and feelings of inadequacy.
Ethics	
Privacy protection	Local, open-source models ensure data security and foster safe exploration.
Academic integrity	Avoids plagiarism by not generating content; instead engages students in reflective dialogue.
Equity	Local deployment may reduce financial and technical barriers, supporting more inclusive access.
Fairness	Socratic questioning aims to reduce exposure to bias/hallucinations and emphasizes critical engagement over information delivery.

3.2 Agency

Regarding agency, Maike adopts two pedagogical principles.

Autonomous Thinking. In addition to fostering critical thinking, Maike is explicitly designed to mitigate the risks of AI-induced dependency, particularly overreliance and the uncritical acceptance of polished, pre-formulated chatbot responses. Maike acts as a partner in critical inquiry instead of positioning itself as an expert, encouraging students to remain cognitively engaged and fostering metacognitive awareness and intellectual autonomy. Therefore, Maike aims to retain the students' agency in their educational journey while benefiting from AI as a reflective scaffold rather than a substitute for thought.

Mitigation of Knowledge Homogenization. Built with a student-centric perspective, Maike supports students to articulate their own ideas, question assumptions, and consider multiple perspectives. This dialectic model is intended to foster intellectual diversity and autonomy, helping students develop original, well-structured positions instead of converging toward standardized or AI-shaped responses.

3.3 Emotion

By engaging students in Socratic dialogue, Maike is intended to support reflection, intellectual autonomy, and sustained cognitive effort without taking over the writing process. We hypothesize

that this form of assistance may support students' sense of ownership and self-efficacy more effectively than answer-giving systems, while future classroom studies are needed to assess whether it also reduces AI-related discomfort, guilt, or feelings of inadequacy across different learners.

3.4 Ethics

In terms of the ethical dimension, Maike focuses on four key elements, described next:

Privacy Protection. Since Maike runs locally on the students' devices using open-source LLMs, it ensures that no personal or academic data is ever sent to external servers. This design protects the students' privacy and empowers institutions, preserving the essential pedagogical value of learning through trial and error, and fosters a safe environment that encourages curiosity and intellectual risk-taking without the pressure of being constantly monitored.

Academic Integrity. Unlike general-purpose AI-based chatbots that often provide direct answers or essay content—thus enabling plagiarism or shortcut-seeking behavior, Maike does not generate content for students. Instead, it engages them in reflective questioning and guided dialogue modeled on Socratic principles, avoiding contributing to academic misconduct.

Equity. Our methodology promotes educational equity by leveraging small, open-source LLMs that run locally on personal computers. This design reduces technical and financial barriers to advanced educational technologies, enabling access regardless of socio-economic background or geographic location. By eliminating reliance on high-end infrastructure or costly cloud services, Maike supports more inclusive and democratized access to personalized learning.

Fairness. Maike is designed to ask targeted, pedagogically informed questions rather than provide direct answers. This Socratic approach minimizes the influence of potential model biases or hallucinations, as the system does not assert content but instead guides students to think critically and develop their own responses. By focusing on dialogue rather than information delivery, the risk of misinformation or uneven guidance is significantly reduced.

4 Current results and future classroom study

We provide an overview of the empirical evaluation of Maike's modules and outline the next steps, including a user study with high school students.

Argument Mining Module. In Favero et al. [8] and Favero et al. [6], we investigate small, open-source decoder-only LLMs for argument mining, focusing on argument component classification and quality assessment in student essays. Contrary to prior work relying on encoder or encoder-decoder models, we show that lightweight LLMs, such as Llama 3.1 8B [2] and Qwen 2.5 7B [14], can be effectively applied via few-shot prompting and fine-tuning. Experiments on the Feedback Prize dataset [1] (grades 6–12) demonstrate that these models outperform baselines in argument type classification and achieve competitive results in quality assessment.

Critical Question Generation Module. In Favero et al. [3], we investigate the potential of LLMs to promote deeper reasoning by generating critical questions that challenge vague or unsupported claims in argumentative discourse. Conducted as part of the shared task on automatic critical question generation at the 12th Workshop on Argument Mining at ACL 2025 [9], this study introduces a two-step framework that leverages two small-scale, open-source LLMs: a Questioner, which generates multiple and diverse candidate questions using various prompts with or without argumentation theory and a Judge, which selects the most relevant and thought-provoking questions. Our system was the winner of the competition, ranking first in the shared task, which illustrates the potential of LLM-based approaches to promote critical engagement with argumentative texts.

Socratic Dialogue Module. In Favero et al. [4], we introduce a first version Maike that implements the Socratic module. This version is based on Paul and Elder [11]'s Socratic method, prompting students to examine diverse perspectives and engage in self-reflection through structured, thought-provoking questions. This approach is achieved by fine-tuning and prompt-tuning an open-source pretrained LLM (we compare Llama 2 models with 7B and 13B parameters [12]) using a specialized dataset designed to encourage critical thinking and explore multiple viewpoints [1]. We evaluate the effectiveness of our approach through a series of experiments simulating student-chatbot interactions, aimed at assessing its impact on students' critical thinking skills. The results show that the Socratic tutor significantly outperforms standard chatbots in fostering reflection and critical reasoning.

Limitations and future work. While preliminary results in evaluating Maike's individual modules are promising, the work remains exploratory. To date, the system has only been evaluated through offline experiments and simulated student interactions, with no classroom deployment. Consequently, important questions remain regarding how real students will experience sustained Socratic friction, including potential frustration, disengagement, or uneven benefits across learners. Moreover, the long-term impact of repeated interaction with Maike on the students' independent essay writing practices and critical thinking habits is still unknown. To address these limitations, we plan to perform in-situ classroom studies to evaluate the full framework with students. These studies will assess the combined contribution of all modules and examine both immediate effects, such as improvements in the current essay, and as medium- and long-term outcomes, including the students' ability to write high-quality essays independently (without relying on Maike) and the development of critical thinking skills over time.

5 Conclusion

This paper presents Maike as a Tool for Thought educational chatbot that prioritizes thinking with AI rather than delegating the thinking to AI. Its main technical contribution is the integration of argument mining, pedagogical strategy selection, critical question generation, and Socratic dialogue within a question-first, local-first design. While current results for the individual modules are promising, this work remains exploratory, and broader claims about

[1] https://www.reddit.com/r/changemyview/

learner agency, emotion, and ethics remain design hypotheses to be tested in future classroom studies.

Acknowledgments

This work has been partially supported by a nominal grant received at the ELLIS Unit Alicante Foundation from the Regional Government of Valencia in Spain (Convenio Singular signed with Generalitat Valenciana, Conselleria de Industria, Turismo, Innovación y Comercio, Dirección General de Innovación). L.F. has also been partially funded by the Bank Sabadell Foundation.

References

[1] Scott Andrew Crossley, Perpetual Baffour, Yu Tian, Alex Franklin, Meg Benner, and Ulrich Boser. 2023. A large-scale corpus for assessing written argumentation: PERSUADE 2.0. *Available at SSRN 4795747* (2023).

[2] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv e-prints* (2024), arXiv–2407.

[3] Lucile Favero, Daniel Frases, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2025. ELLIS Alicante at CQs-Gen 2025: Winning the critical thinking questions shared task: LLM-based question generation and selection. *arXiv preprint arXiv:2506.14371* (2025).

[4] Lucile Favero, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2024. Enhancing critical thinking in education by means of a Socratic chatbot. In *International Workshop on AI in Education and Educational Research*. Springer, 17–32.

[5] Lucile Favero, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2024. Towards Student-Centric AI-Supported Learning: Teaching Chatbots to Ask the Right Questions. *Collaborative AI and Modeling of Humans, AAAI Bridge Program* (2024).

[6] Lucile Favero, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2025. Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. *arXiv preprint arXiv:2502.14389* (2025).

[7] Lucile Favero, Juan Antonio Pérez-Ortiz, Tanja Käser, and Nuria Oliver. 2026. AI in Education Beyond Learning Outcomes: Cognition, Agency, Emotion, and Ethics. *arXiv preprint arXiv:2602.04598* (2026).

[8] Lucile Favero, Juan Antonio Pérez-Ortiz, and Tanja Käser Nuria Oliver. 2025. Argument Mining in Education: Exploring the Potential of Open-source Small LLMs for Argument Classification and Assessment. (2025).

[9] Blanca Figueras, Jaione Bengoetxea, Maite Heredia, Ekaterina Sviridova, Elena Cabrio, Serena Villata, and Rodrigo Agerri. 2025. Overview of the Critical Questions Generation Shared Task 2025. In *Proceedings of the 12th Workshop on Argument Mining*.

[10] Michael Gerlich. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *MDPI* (2025).

[11] Richard Paul and Linda Elder. 2007. Critical thinking: The art of Socratic questioning. *Journal of developmental education* 31, 1 (2007), 36.

[12] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[13] Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 176–187.

[14] An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115* (2024).

[15] Meriem Zerkouk, Miloud Mihoubi, and Belkacem Chikhaoui. 2025. A Comprehensive Review of AI-based Intelligent Tutoring Systems: Applications and Challenges. *arXiv preprint arXiv:2507.18882* (2025).