

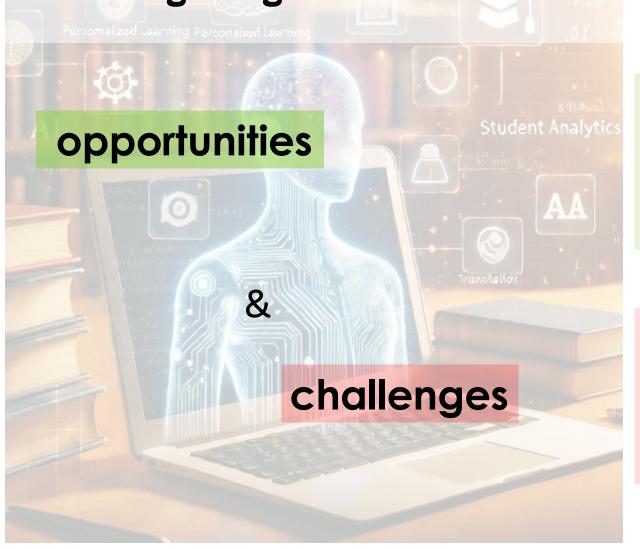








Increasing integration of AI tools in education presents both



- Mitigate teacher shortages
- Up-to-date educational materials
- Self-paced learning
- 24/7 Instant feedback
- Risk of overreliance
- Critical thinking
- student agency
- Concerns about data privacy

Towards a Critical &

Responsible Use of Al in Education

COGNITION

Cognitive offloading

Superficial learning



Desirable difficulty
Pedagogical principles
Stimulation of critical thinking

ETHICS

Surveillance & privacy loss

Low academic integrity

Data exploitation



Strong ethical frameworks & policies

EMOTION

Al guilt/stress/anxiety

Cognitive dissonance

Low self-esteem/self-efficacy



Al demystification Meaningful human interaction

AGENCY

Dependency & overtrust Knowledge conformity



Reflection on AI use Purposeful AI use AI literacy

[1] Favero, Perez-Ortiz, Kaser, Oliver. Do AI tutors empower or enslave learners? Toward a critical use of AI in education GenAIHE workshop @ AIED2025

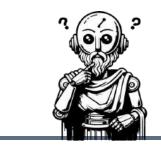


Rather than allowing chatbots to weaken our cognitive abilities, my objective is to develop a chatbot that supports the enhancement of critical thinking and self-regulation skills

A Socratic approach to an educational chatbot



Provides the knowledge to facilitate the learning



Socratic chatbot

Asks specific and targeted questions to help the learners build knowledge by themselves

Socratic questioning

What is Socratic questioning?

Socratic questioning is a **disciplined questioning** method used to **encourage critical thinking**, uncover underlying beliefs, and explore complex ideas

Socratic questioning

Taxonomy of Socratic questions based on Critical Thinking concepts¹

Question type

Description

Exemplar

01	Clarification	Question probing the ambiguities of a thought	What do you mean by ?
02	Probing assumption	Question probing the assumptions behind a thought	Why do you assume ?
03	Probing reasons and evidences	Question probing the justifications or concrete evidences that could have supported a thought	How did you know that ?
04	Probing implications and consequences	Question probing the impacts or implications of a thought	If, what is likely to happen as a result?
05	Probing alternative viewpoints and perspectives	Question probing other possible viewpoints	What else should we consider about?

^{1.} R. Paul, L. Elder, The Thinker's Guide to Socratic Questioning, Rowman & Littlefield, 2019.

A Socratic approach to an educational chatbot

Aims

- Reflexion
- Purposeful learning
- Critical thinking

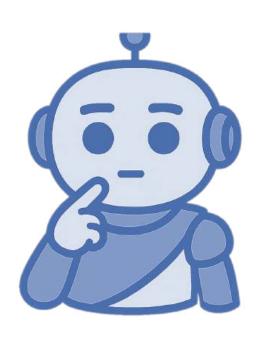
Broader applications

- Combatting fake news
- Promoting democracy

Educational Tasks

- Question answering
- Self-regulation
- Knowledge discovery

Focus on helping students improve their essays



The educational chatbot to promote critical thinking

1. The student writes their text.

Hi Lucile, ready to write? Please enter your essay details

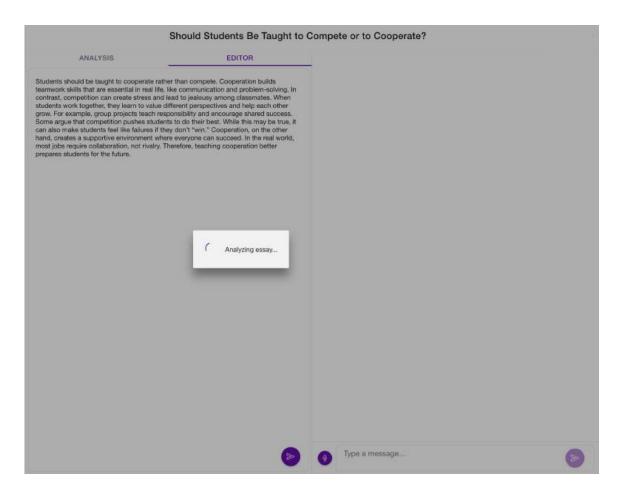
Should Students Be Taught to Compete or to Cooperate?

Students should be taught to cooperate rather than compete. Cooperation builds teamwork skills that are essential in real life, like communication and problem-solving. In contrast, competition can create stress and lead to jealousy among classmates. When students work together, they learn to value different perspectives and help each other grow. For example, group projects teach responsibility and encourage shared success. Some argue that competition pushes students to do their best. While this may be true, it can also make students feel like failures if they don't "win." Cooperation, on the other hand, creates a supportive environment where everyone can succeed.

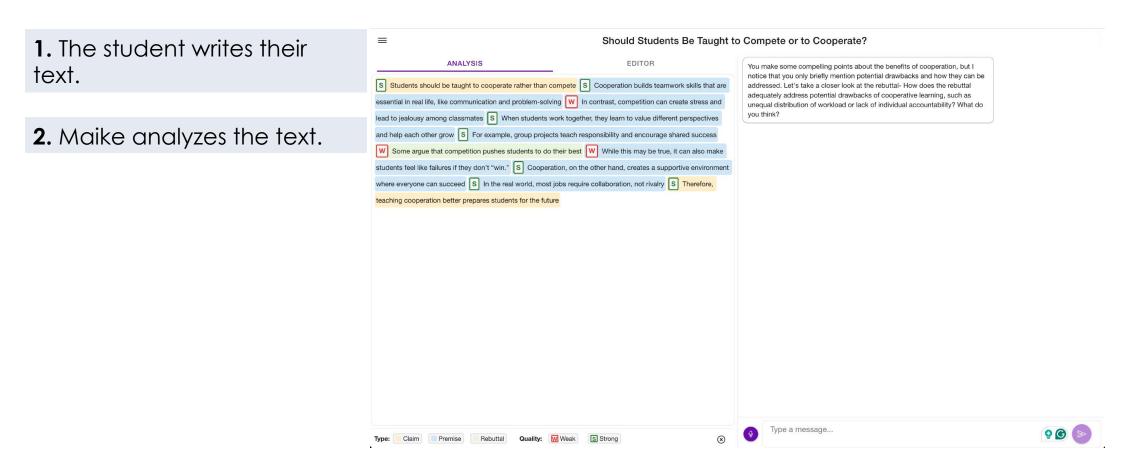
>

1. The student writes their text.

2. Maike analyzes the text.



Frontend & UI: Marta Serrador



Frontend & UI: Marta Serrador

1. The student writes their text.

 \equiv

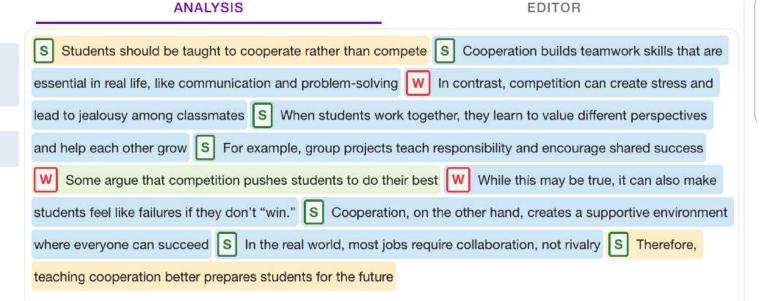
- 2. Maike analyzes the text.
- Detection of argument components

Claim Premise Rebuttal

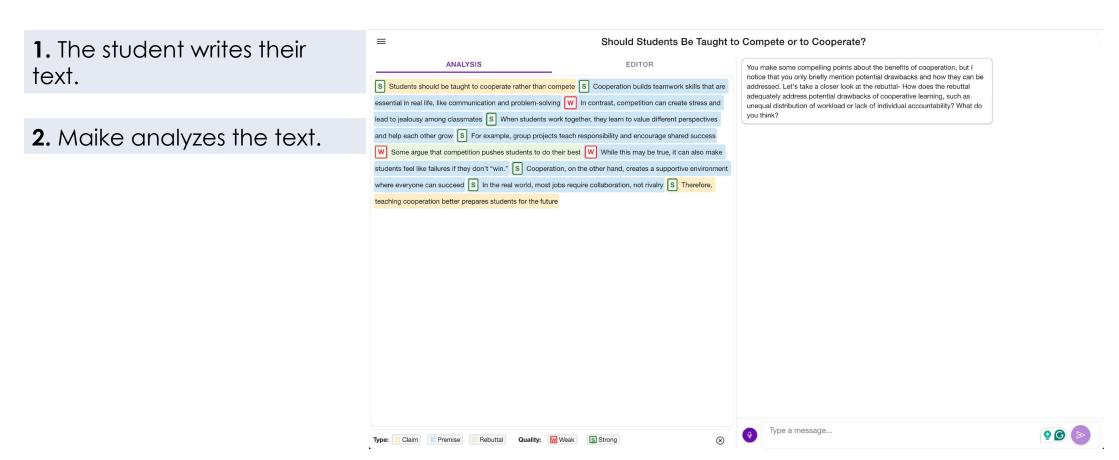
Quality assessment of each component



Should Students Be Taught to



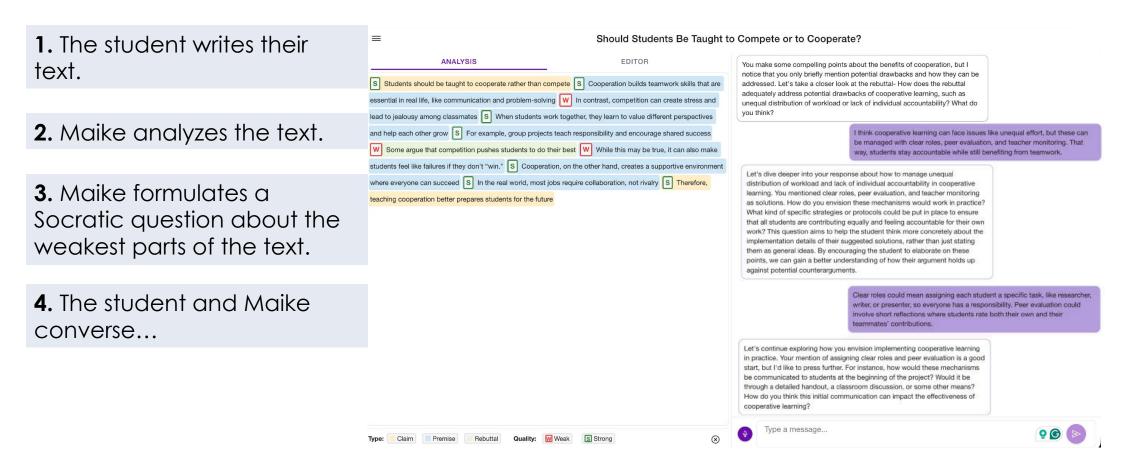
Frontend & UI: Marta Serrador



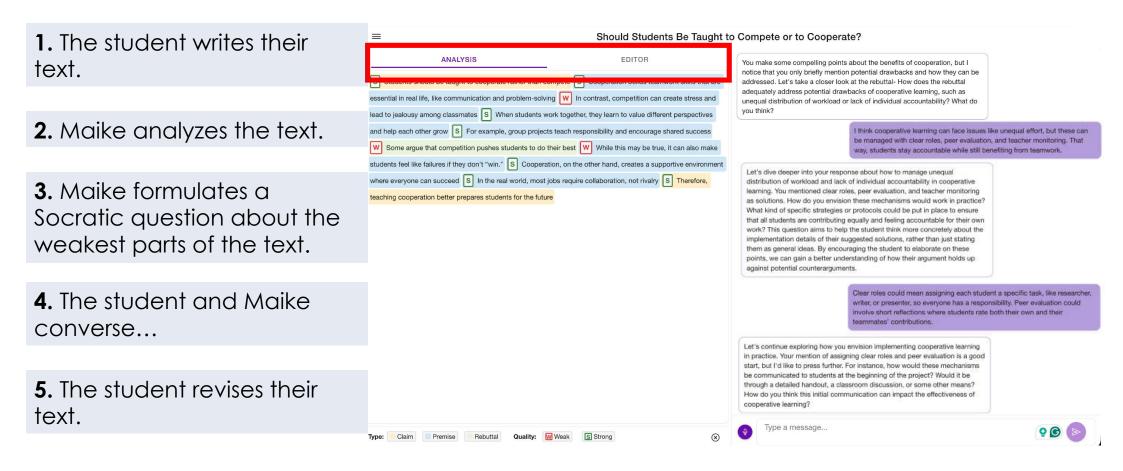
Frontend & UI: Marta Serrador

- 1. The student writes their text.
- 2. Maike analyzes the text.
- **3.** Maike formulates a Socratic question about the weakest parts of the text.

You make some compelling points about the benefits of cooperation, but I notice that you only briefly mention potential drawbacks and how they can be addressed. Let's take a closer look at the rebuttal—How does the rebuttal adequately address potential drawbacks of cooperative learning, such as unequal distribution of workload or lack of individual accountability? What do you think?

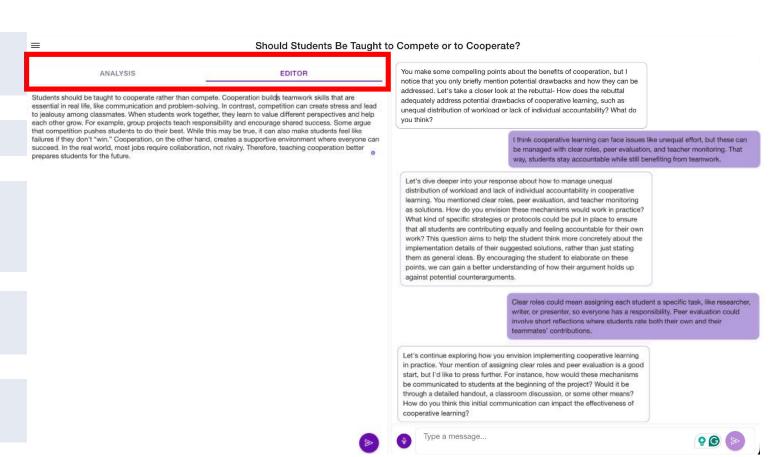


Frontend & UI: Marta Serrador



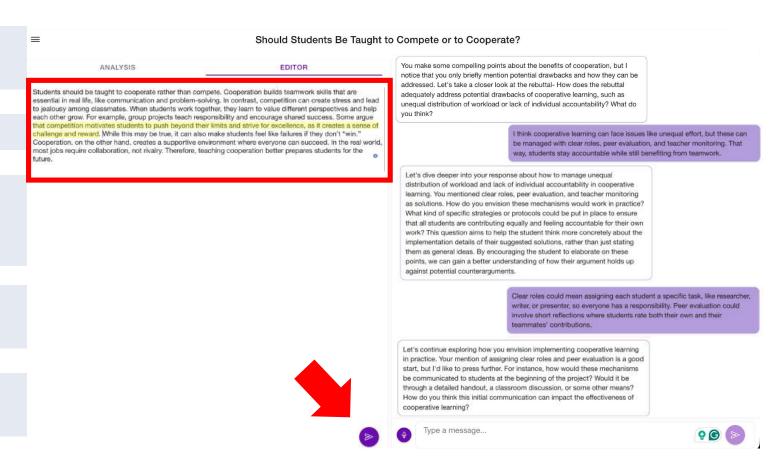
Frontend & UI: Marta Serrador

- 1. The student writes their text.
- 2. Maike analyzes the text.
- **3.** Maike formulates a Socratic question about the weakest parts of the text.
- **4.** The student and Maike converse...
- **5.** The student revises their text.

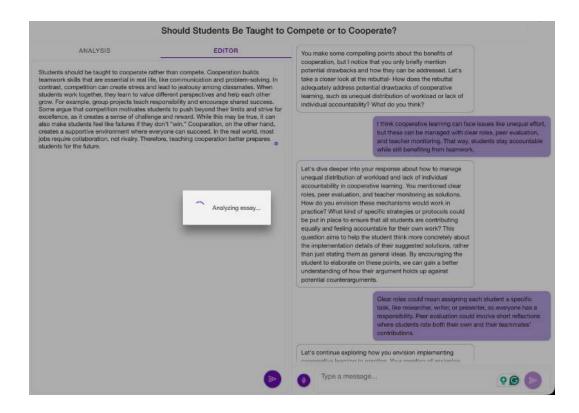


Frontend & UI: Marta Serrador

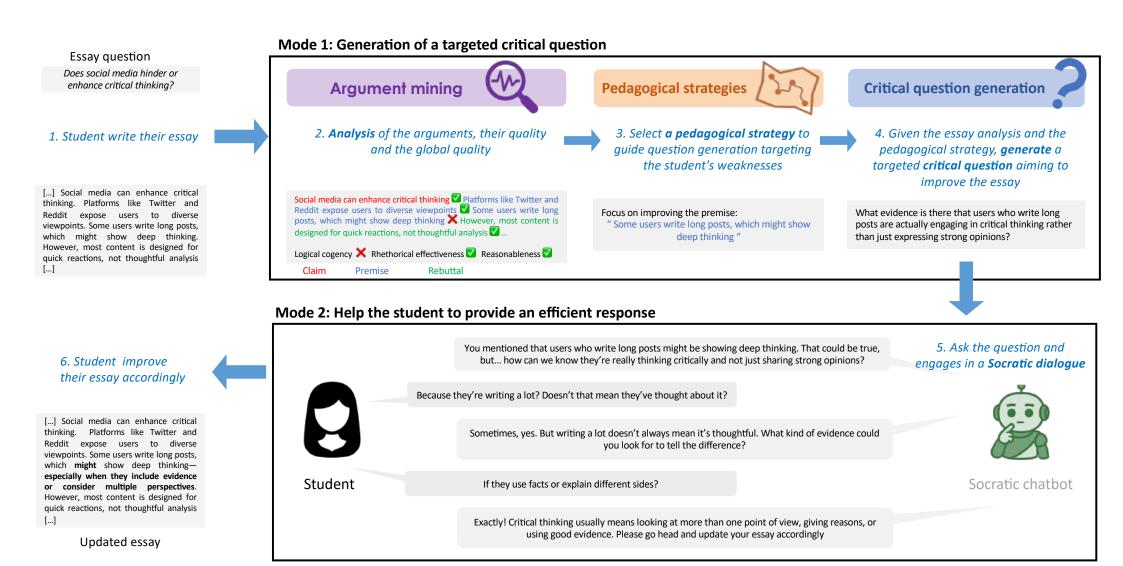
- 1. The student writes their text.
- 2. Maike analyzes the text.
- **3.** Maike formulates a Socratic question about the weakest parts of the text.
- **4.** The student and Maike converse...
- **5.** The student revises their text.



- 1. The student writes their text.
- 2. Maike analyzes the text.
- **3.** Maike formulates a Socratic question about the weakest parts of the text.
- **4.** The student and Maike converse...
- **5.** The student revises their text.
- **6.** Maike analyzes the new text.



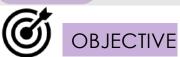
Frontend & UI: Marta Serrador



[2] Favero, Perez-Ortiz, Kaser, Oliver. <u>Towards Student-Centric AI-Supported Learning: Teaching Chatbots</u> to Ask the Right Questions. Bridge Program Collaborative AI and Modeling of Humans @ AAAI2024



Leveraging small LLMs for argument mining in education: Argument component identification, classification, and Assessment

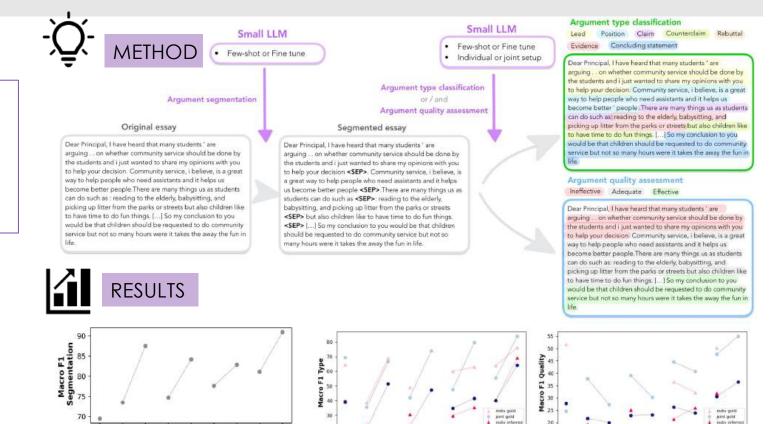


Performing **argument mining** with **small open - source** LLMs:

- Component segmentation
- Type classification
- Quality assessment



- Dataset: Feedback Prize
- Metric: Macro-averaged F1
- Baseline: BERT (two heads) and Longformer



[5] Favero, Pérez-Ortiz, Käser, & Oliver. Argument Mining in Education: Exploring the Potential of Open-source Small LLMs for Argument Classification and Assessment. AI for Education Tools, Opportunities, and Risks in the Generative AI Era @ AAAI2025 and the 25th International Workshop on Computational Models of Natural Argument



Winning the critical thinking questions shared task: LLM-based question generation and selection



The dataset Interventions of real debates, annotated with Argumentation Schemes and a set of Critical Questions and their usefulness for challenging the arguments: Useful, Unhelpful, or Invalid

The task Generate three critical auestions that are Useful

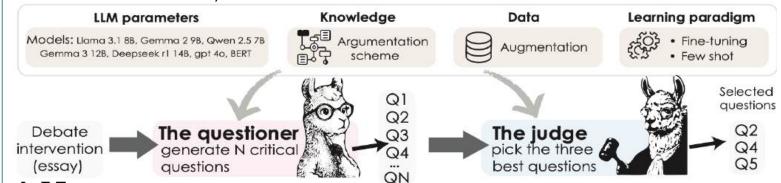


EVALUATION

- Compute cosine similarity of the question with the annotated references
- Limitations of automatic assessment in such a complex and open task



METHOD Questioner-Judge LLM architecture based on critical thinking theory: analytic, creative, and evaluative dimensions





Questioner	Scheme in the prompt			% Useful	% Un- matched	
Gemma 3 12B	without	8	π	40.5	46.0	
Llama FT ***	without	8	9	56.6	29.4	
Llama 3.1 8B	without	8	Deepseek 14B	57.4	27.0	
Gemma 2 9B	with	8	Llama 3.1 8B	54.7	31.6	
Llama 3.1 8B	with	6	Gemma 3 12B	49.4	34.1	
Llama 3.1 8B	both	12	Gemma 2 9B	62.4	25.7	
GPT 40	both	12	GPT 40	62.7	25.0	

***Llama 3.1 8B fine-tuned

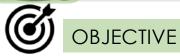
Best configuration

- Questioner + Judae
- With and without schemes
- Small open-source LLMs: Llama 3.1 8B, Gemma 2 9B
- Chat GPT 40

[4] Favero, Frases, Pérez-Ortiz, Käser, & Oliver ELLIS Alicante at CQs-Gen 2025: Winning the critical thinking questions shared task: LLM-based question generation and selection. 12th Workshop on Argument Mining @ACL 2025.



Enhancing critical thinking in education by means of a Socratic chatbot



Building a Socratic chatbot:

Generation of Socratic questions by prompt and fine-tuning a pre-trained LLM



1. **SocratiQ:** Dataset of annotated Socratic questions and answers using r/changemyview



2. Fine-tuning Llama2 7B and 13B four-bit quantized with unsloth



3. Prompt-tuning of the finetuned model on Ollama





Compare the learner's response and the ground truth

- Generation of synthetic conversations: prompt-tuning Llama 2 7B to simulate learner responses
- Baselines: 2 non-Socratic prompt-tuned tutors
- Evaluation Dataset: Theory of Knowledge International Baccalaureate questions



RESULTS

The Socratic tutors outperform the baselines on all metrics

utor type	BLEU	ROUGE-	METEOR	BERTSco	re LLM						
		L			Score	0.25	Type of Convers	ation M	(ETEO	R Sco	re
Socratic Llama2 13B	3.65	0.157	0.226	0.569	0.696	0.20	Basic tutor Socratic tuto	r 78		-	75.00
Socratic Llama2 7B	3.42	0.162	0.216	0.576	0.670	METEOR Score					
Basic Llama2 7B	0.494	0.120	0.092	0.535	0.582	0.05					
Random Llama2 7B	0.210	0.091	0.063	0.444	0.312	0.00	0	i	ž Tum nu	3 mber	

[3] Favero, Pérez-Ortiz, Käser, & Oliver <u>Enhancing Critical Thinking in Education by means of a Socratic Chatbot</u>.

International Workshop on AI in Education and Educational Research @ ECAI 2024.



Example of conversation with Maike

Is replicability necessary in the production of knowledge?

Clarification

What do you mean by "replicability"?



Reasons

Replicability means being able to reproduce or verify the results of an experiment or study by repeating it with similar methods and obtaining similar results.

> Why do you think you need replicability in order to produce knowledge?



Replicability is necessary for producing knowledge because it allows us to verify and confirm the accuracy of findings through independent repetition of experiments or observations.

Alternative viewpoints

How would you respond to someone who argues that replication is not necessary in this case because the sample size was so small?



Next steps

App development

Building a conversational interface as a functional software tool

Chatbot improvement

Reinforcement learner: construction of an optimal learning path

Personalization to improve student engagement and outcomes

User study

International Bacalaureat

At EPFL, Prof. Tanja Käser







Collaboration

Automated Essay Scoring





Thank you for your attention

Project presentation: https://ellisalicante.org/maike



- Posters:
 - https://web.ua.es/de/phdinf/documentos/jdi-2024/lucilealysfavero.pdf
 - https://web.ua.es/es/phdinf/documentos/jdi-2025/lucilealysfaveromontero.pdf
- Scientific papers: https://ellisalicante.org/people/lucilefavero-en/
- **Talk:** https://web.ua.es/en/iuii/noticias/talk-lucile-favero-towards-student-centric-ai-supported-learning-teaching-chatbots-to-ask-the-right-questions.html
- Contact: <u>lucile@ellisalicante.org</u> <u>http://lucilefavero.github.io/</u>













Publications

2025	
12/12	Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2025). Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. The 25th International Workshop on Computational Models of Natural Argument (CMNA'25).
07/31 Vienna, AT	Favero, L. A., Frases, D., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2025). <u>ELLIS Alicante at CQs-Gen 2025</u> : Winning the critical thinking questions shared task: <u>LLM-based question generation and selection</u> . <i>12th Workshop on Argument Mining, co-located with ACL 2025</i> .
07/22 Palermo, IT	Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2025). Do Al tutors empower or enslave learners? Toward a critical use of Al in education. Applications of generative Al to support teaching and learning in higher education, co-located with AIED 2025.
03/03 Philadelphia, US	Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2025). Argument Mining in Education: Exploring the Potential of Open-source Small LLMs for Argument Classification and Assessment. AAAI2025 AI for Education - Tools, Opportunities, and Risks in the Generative AI Era.
02/20	Favero, L. A. , Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2025). <u>Leveraging Small LLMs for Argument Mining in Education: Argument Component Identification, Classification, and Assessment. <i>arXiv:2502.14389</i>.</u>
2024	
10/19 Santiago de Compostela, ES	Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024). Enhancing Critical Thinking in Education by means of a Socratic Chatbot. International Workshop on Al in Education and Educational Research, ECAI 2024 workshop.
02/21	Favero, L. A., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024). Towards Student-Centric Al- Supported Learning: Teaching Chatbots to Ask the Right Questions. Collaborative Al and Modeling of Humans, AAAI Bridge Program.