An LLM-based educational chatbot to promote Critical Thinking

Lucile Favero Montero

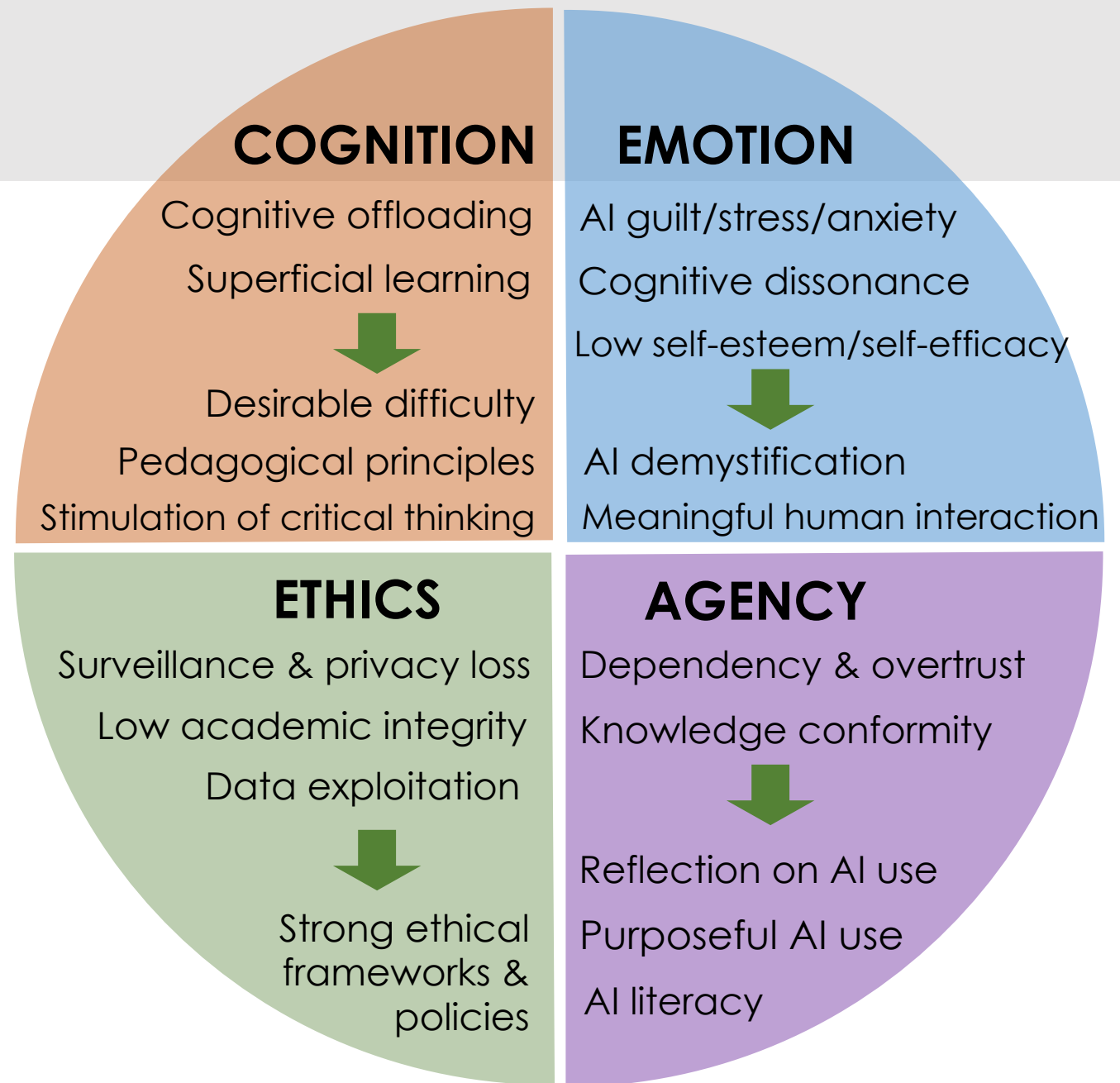# Increasing integation of AI tools in education presents both

opportunities

&

challenges

- 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 AI in Education

COGNITION
- Cognitive offloading
- Superficial learning

↓

- Desirable difficulty
- Pedagogical principles
- Stimulation of critical thinking

EMOTION
- AI guilt/stress/anxiety
- Cognitive dissonance
- Low self-esteem/self-efficacy

↓

- AI demystification
- Meaningful human interaction

ETHICS
- Surveillance & privacy loss
- Low academic integrity
- Data exploitation

↓

- Strong ethical frameworks & policies

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

Traditional 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 concepts[1]

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

Educational Tasks

- Question answering
- Self-regulation
- Knowledge discovery

Broader applications

- Combatting fake news
- Promoting democracy

Focus on helping students improve their essays

Mode 1: Generation of a targeted critical question

Essay question

Does social media hinder or enhance critical thinking?

1. Student write their essay

[...] Social media can enhance critical thinking. Platforms like Twitter and Reddit expose users to diverse viewpoints. Some users write long posts, which might show deep thinking. However, most content is designed for quick reactions, not thoughtful analysis [...]

Argument mining

2. **Analysis** of the arguments, their quality and the global quality

Social media can enhance critical thinking ✅ Platforms like Twitter and Reddit expose users to diverse viewpoints ✅ Some users write long posts, which might show deep thinking ❌ However, most content is designed for quick reactions, not thoughtful analysis ✅ ...

Logical cogency ❌ Rhethorical effectiveness ✅ Reasonableness ✅

Claim   Premise   Rebuttal

Pedagogical strategies

3. Select **a pedagogical strategy** to guide question generation targeting the student's weaknesses

Focus on improving the premise:
" Some users write long posts, which might show deep thinking "

Critical question generation

4. Given the essay analysis and the pedagogical strategy, **generate** a targeted **critical question** aiming to improve the essay

What evidence is there that users who write long posts are actually engaging in critical thinking rather than just expressing strong opinions?

Mode 2: Help the student to provide an efficient response

6. Student improve their essay accordingly

[...] Social media can enhance critical thinking. Platforms like Twitter and Reddit expose users to diverse viewpoints. Some users write long posts, which **might** show deep thinking— **especially when they include evidence or consider multiple perspectives**. However, most content is designed for quick reactions, not thoughtful analysis [...]

Updated essay

Student

You mentioned that users who write long posts might be showing deep thinking. That could be true, but... how can we know they're really thinking critically and not just sharing strong opinions?
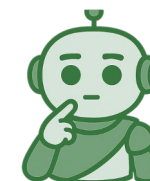
Because they're writing a lot? Doesn't that mean they've thought about it?

Sometimes, yes. But writing a lot doesn't always mean it's thoughtful. What kind of evidence could you look for to tell the difference?

If they use facts or explain different sides?

Exactly! Critical thinking usually means looking at more than one point of view, giving reasons, or using good evidence. Please go head and update your essay accordingly

5. Ask the question and engages in a **Socratic dialogue**

Socratic chatbot

[2] Favero, Perez-Ortiz, Kaser, Oliver. Towards Student-Centric AI-Supported Learning: Teaching Chatbots 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

OBJECTIVE

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

- Component segmentation
- Type classification
- Quality assessment

EVALUATION

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

METHOD

Small LLM
- Few-shot or Fine tune

Small LLM
- Few-shot or Fine tune
- Individual or joint setup

Argument segmentation

Argument type classification
or / and
Argument quality assessment

Original essay

Dear Principal, I have heard that many students ' are arguing . . on whether community service should be done by the students and i just wanted to share my opinions with you to help your decision. Community service, i believe, is a great way to help people who need assistants and it helps us become better people.There are many things us as students can do such as : reading to the elderly, babysitting, and picking up litter from the parks or streets but also children like to have time to do fun things. […] So my conclusion to you would be that children should be requested to do community service but not so many hours were it takes the away the fun in life.

Segmented essay

Dear Principal, I have heard that many students ' are arguing . . on whether community service should be done by the students and i just wanted to share my opinions with you to help your decision **<SEP>**. Community service, i believe, is a great way to help people who need assistants and it helps us become better people **<SEP>**.There are many things us as students can do such as **<SEP>**: reading to the elderly, babysitting, and picking up litter from the parks or streets **<SEP>** but also children like to have time to do fun things. **<SEP>** […] So my conclusion to you would be that children should be requested to do community service but not so many hours were it takes the away the fun in life.
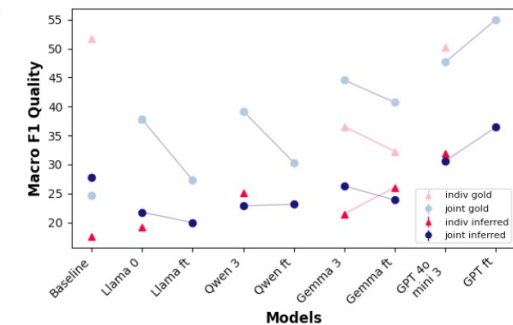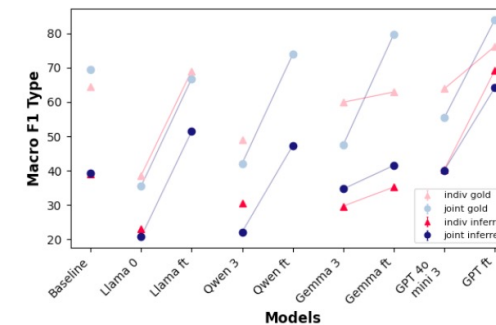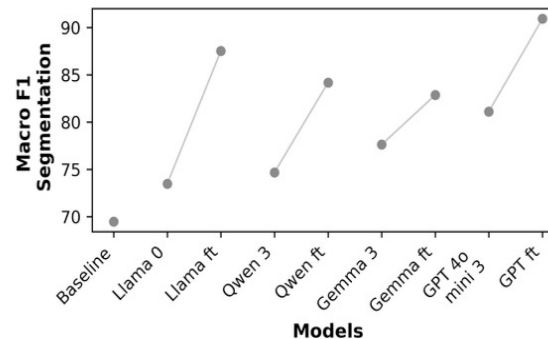
Argument type classification

Lead Position Claim Counterclaim Rebuttal
Evidence Concluding statement

Dear Principal, I have heard that many students ' are arguing . . on whether community service should be done by the students and i just wanted to share my opinions with you to help your decision. Community service, i believe, is a great way to help people who need assistants and it helps us become better ' people .There are many things us as students can do such as: reading to the elderly, babysitting, and picking up litter from the parks or streets but also children like to have time to do fun things. […] So my conclusion to you would be that children should be requested to do community service but not so many hours were it takes the away the fun in life.

Argument quality assessment

Ineffective Adequate Effective

Dear Principal, I have heard that many students ' are arguing . . on whether community service should be done by the students and i just wanted to share my opinions with you to help your decision. Community service, i believe, is a great way to help people who need assistants and it helps us become better people.There are many things us as students can do such as: reading to the elderly, babysitting, and picking up litter from the parks or streets but also children like to have time to do fun things. […] So my conclusion to you would be that children should be requested to do community service but not so many hours were it takes the away the fun in life.

RESULTS

[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*

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

OBJECTIVE

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

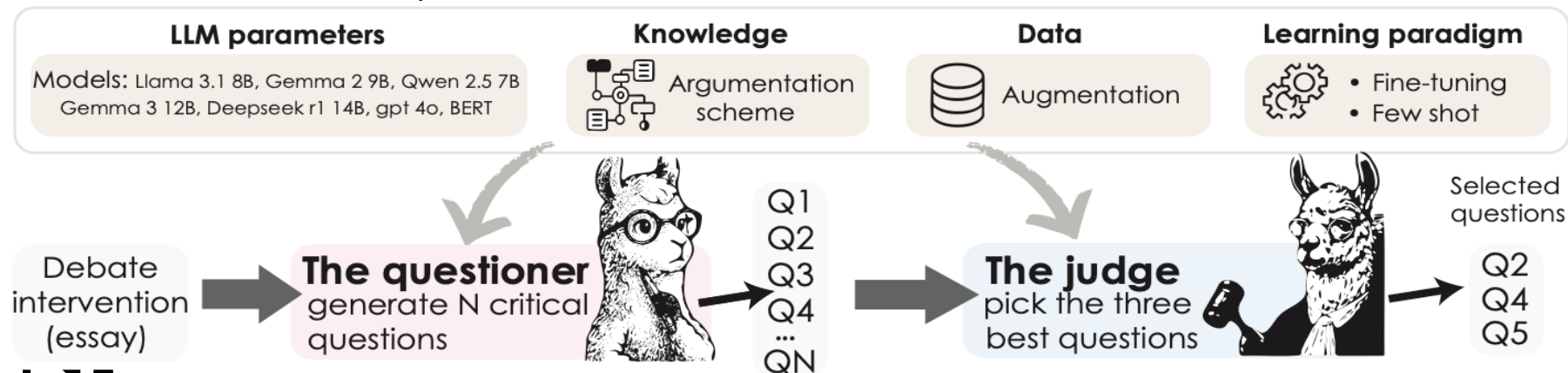**The task** Generate three critical questions that are Useful

METHOD

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

RESULTS

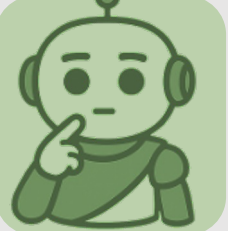| Questioner	Scheme in the prompt	#Questions generated	Judge	% Useful	% Un-matched
Gemma 3 12B	without	8	-	40.5	46.0
Llama FT ***	without	8	-	56.6	29.4
Llama 3.1 8B	without	8	Deepseek 14B	57.4	27.0
Gemma 2 9B	with	8	Llama 3.1 8B	54.7	31.6
Llama 3.1 8B	with	6	Gemma 3 12B	49.4	34.1
Llama 3.1 8B	both	12	Gemma 2 9B	**62.4**	**25.7**
GPT 4o	both	12	GPT 4o	**62.7**	**25.0**

***Llama 3.1 8B fine-tuned

EVALUATION

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

Best configuration

- Questioner + Judge
- With and without schemes
- Small open-source LLMs:
  Llama 3.1 8B, Gemma 2 9B
- Chat GPT 4o

[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

OBJECTIVE

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

METHOD

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 fine-tuned model on Ollama
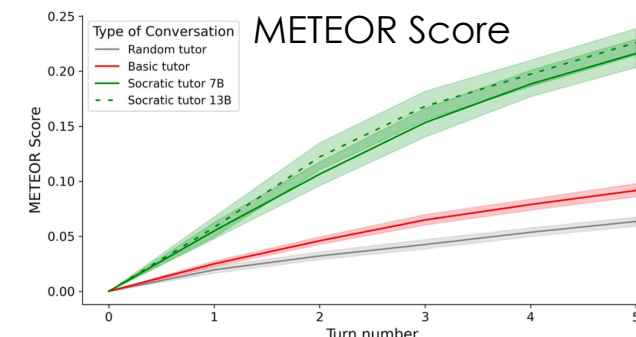
EVALUATION

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

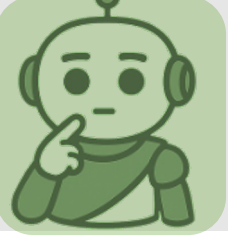- *Evaluation Dataset*: Theory of Knowledge International Baccalaureate questions

RESULTS

The Socratic tutors outperform the baselines on all metrics

Tutor type	BLEU	ROUGE-L	METEOR	BERTScore	LLM Score
Socratic Llama2 13B	**3.65**	**0.157**	**0.226**	**0.569**	**0.696**
Socratic Llama2 7B	**3.42**	**0.162**	**0.216**	**0.576**	**0.670**
Basic Llama2 7B	0.494	0.120	0.092	0.535	0.582
Random Llama2 7B	0.210	0.091	0.063	0.444	0.312

METEOR Score

[3] Favero, Pérez-Ortiz, Käser, & Oliver Enhancing Critical Thinking in Education by means of a Socratic Chatbot. *International Workshop on AI in Education and Educational Research @ ECAI 2024.*

Next steps

App and Webapp

Marta Serrador, Nuria Riera, Daniel Frases

Universitat d'Alacant
Universidad de Alicante

UAX Universidad Alfonso X el Sabio

Collaboration

Argument mining
Prof Ramón Ruiz-Dolz

University of Dundee ARG-tech

User study

- International Bacalaureat

SEK EDUCATION GROUP

- At EPFL, Prof. Tanja Käser

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Chatbot improvement

Reinforcement learner: construction of an optimal learning path

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:** lucile@ellisalicante.org   http://lucilefavero.github.io/