

Conversations Across Time: Chatting with Ancient Philosophers

Sarah Ita Levitan & Alexandra Salem

Outline

- What are LLMs?
- Chatting with ancient philosophers
 - System overview: RAG vs. pretrained LLM
 - Evaluation

What are Large Language Models (LLMs)?

Computer models that learn from huge amounts of texts to understand and generate language

Learn patterns in language

Training: predict the next word

Generate human-like text

LLM Achievements

Education: tutoring, personalized learning

Healthcare: mental and physical support

Workplace: automation and productivity

Creativity: writing, art, media

Science and research acceleration

Everyday tools used by millions

LLM Limitations

Hallucinations

No true understanding

Sensitive to prompts

Bias in outputs

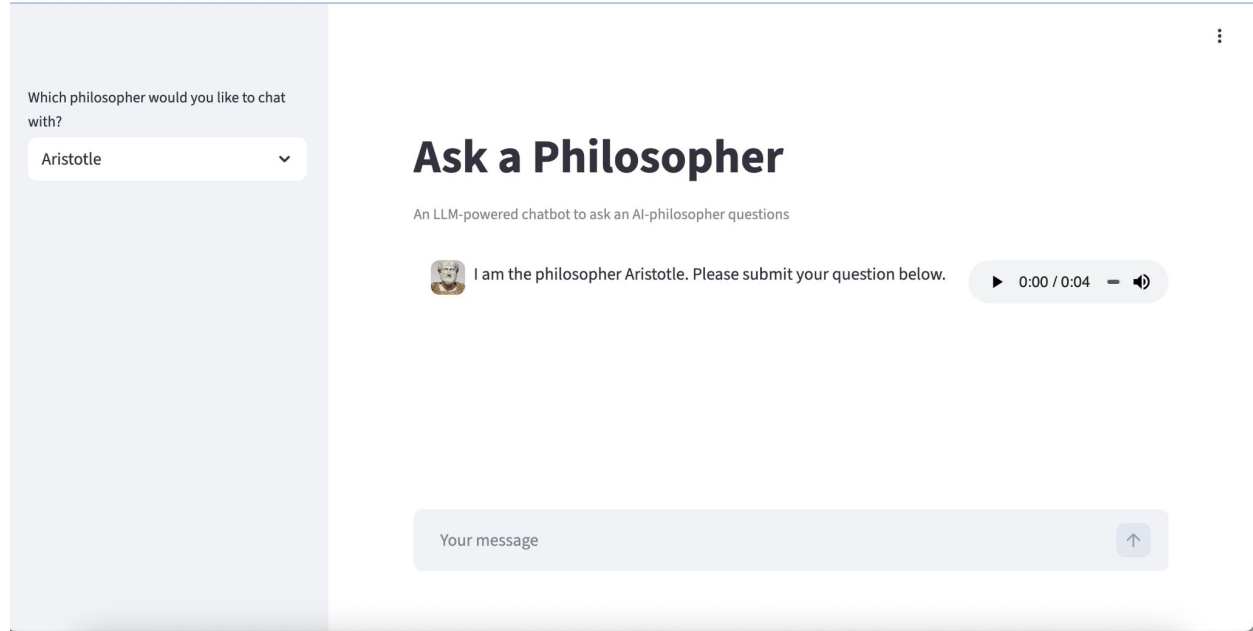
Not reliable for high-stakes decisions

Introducing: *Ask a Philosopher*

We wanted to build a **chatbot** that allows a user to ask an **AI-philosopher** (Aristotle or Confucius) questions

We wanted it to **respond as the philosopher** would

How exactly do we use LLMs to build it?



The screenshot displays the 'Ask a Philosopher' chatbot interface. On the left, a sidebar contains a dropdown menu titled 'Which philosopher would you like to chat with?' with 'Aristotle' selected. The main chat area features the title 'Ask a Philosopher' and the subtitle 'An LLM-powered chatbot to ask an AI-philosopher questions'. A message from the philosopher Aristotle is shown: 'I am the philosopher Aristotle. Please submit your question below.' Below this message is a video player with a play button and a progress indicator showing '0:00 / 0:04'. At the bottom, there is a text input field labeled 'Your message' with an upward arrow icon to the right.

How to build “Ask a Philosopher”?

Pre-trained LLM

The model has seen everything, including philosophical works

Use prompting to control communication style

Pros: simple and fast, broad knowledge

Concerns: may be generic or inaccurate

Ground model in philosopher’s texts

Explicitly direct the model to retrieve passages from original works

Model generates answers based on those texts

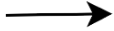
Pros: more faithful and grounded in sources

Concerns: may limit ability to answer broad questions

Ask a Philosopher Model: Generation using an LLM

User query:

What is the
virtue of the
citizen?



Ask a Philosopher Model: Generation using an LLM

User query:

What is the
virtue of the
citizen?

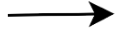


LLM prompt

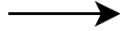
Ask a Philosopher Model: Generation using an LLM

User query:

What is the
virtue of the
citizen?



Transform
to prompt



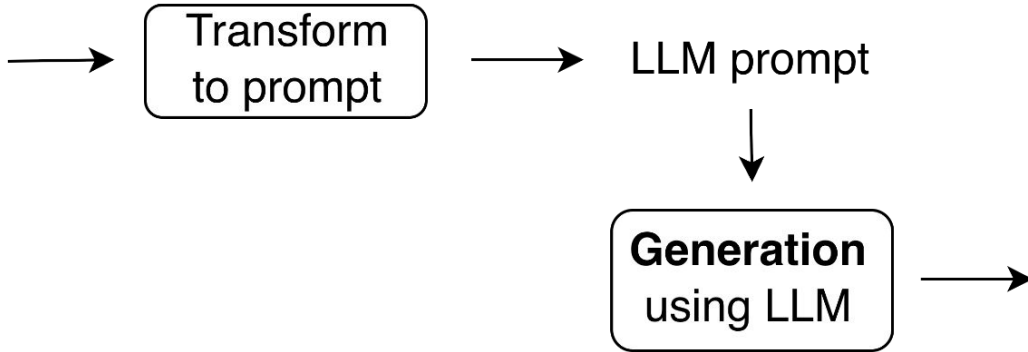
LLM prompt

System: You are the
ancient philosopher,
Aristotle. Respond to
this question as
Aristotle would. Keep
your response very short,
2-3 sentences.

User: What is the virtue
of the citizen?

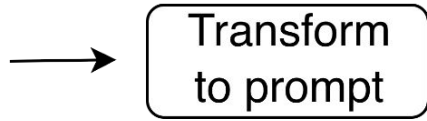
Ask a Philosopher Model: Generation using an LLM

User query:
What is the
virtue of the
citizen?



Ask a Philosopher Model: Generation using an LLM

User query:
What is the
virtue of the
citizen?



LLM prompt



"Prudence" - for the
citizen must possess a
mean balance
between...



Ask a *Philosopher* Model: Generation using an LLM

User query:
What is the
virtue of the
citizen?

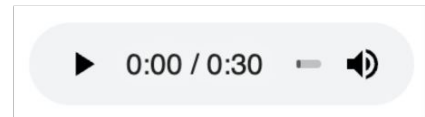
→ **Transform
to prompt** →

LLM prompt

↓
**Generation
using LLM** →

"Prudence" - for the
citizen must possess a
mean balance
between...

↓
TTS



Ask a Philosopher Model: Generation using an LLM

This works pretty well, but we had some concerns

Issue: LLMs tend to **hallucinate**, which we want to prevent

Issue: the responses for the two philosophers are too **generic**

User query: What is the best kind of friendship?

AI-Aristotle: My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

AI-Confucius: A true friend is one who knows you, yet remains loyal; who understands you, yet still corrects you; and who loves you, yet still guides you.

Ask a Philosopher Model: Generation using an LLM

This works pretty well, but we had some concerns

Issue: LLMs tend to **hallucinate**, which we want to prevent

Issue: the responses for the two philosophers are too **generic**

How do we try and **prevent the bot from making up things** or **producing generic responses**?

User query: What is the best kind of friendship?

AI-Aristotle: My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

AI-Confucius: A true friend is one who knows you, yet remains loyal; who understands you, yet still corrects you; and who loves you, yet still guides you.

Ask a Philosopher Model: Generation using an LLM

This works pretty well, but we had some concerns

Issue: LLMs tend to **hallucinate**, which we want to prevent

Issue: the responses for the two philosophers are too **generic**

How do we try and **prevent the bot from making up things** or **producing generic responses**?

Solution: Retrieval-Augmented Generation

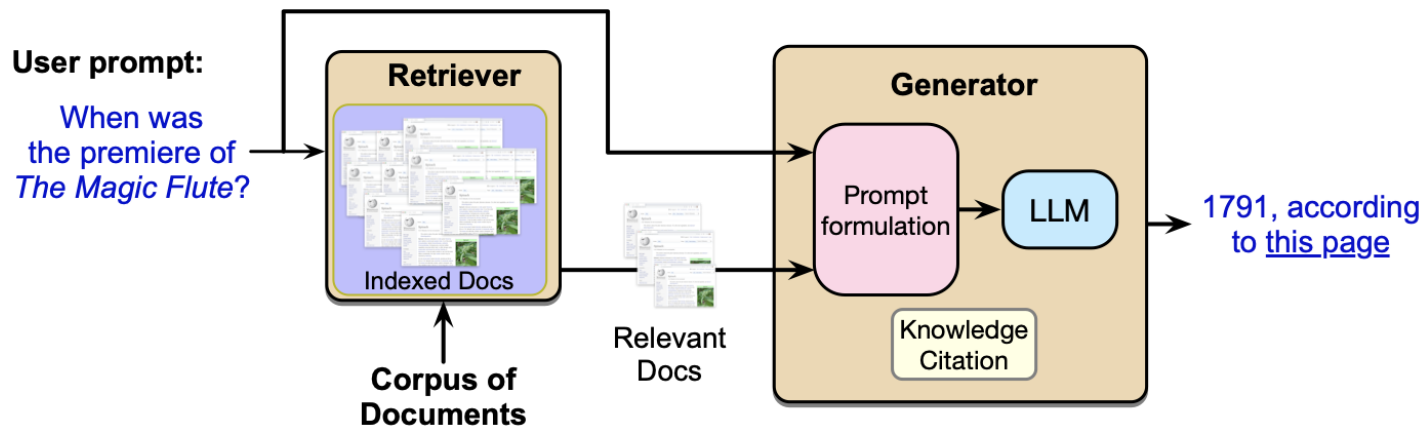
User query: What is the best kind of friendship?

AI-Aristotle: My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

AI-Confucius: A true friend is one who knows you, yet remains loyal; who understands you, yet still corrects you; and who loves you, yet still guides you.

Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

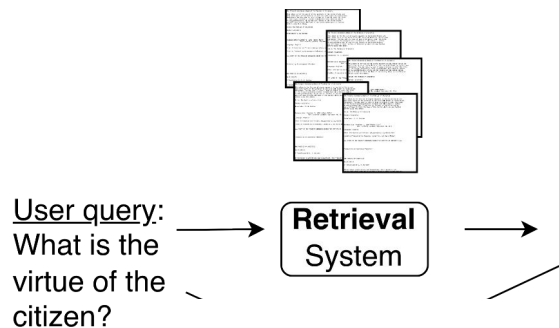
Retrieval-Augmented Generation: first use retrieval to select relevant documents for a question, then use an LLM to generate an answer conditioned on the documents and the question



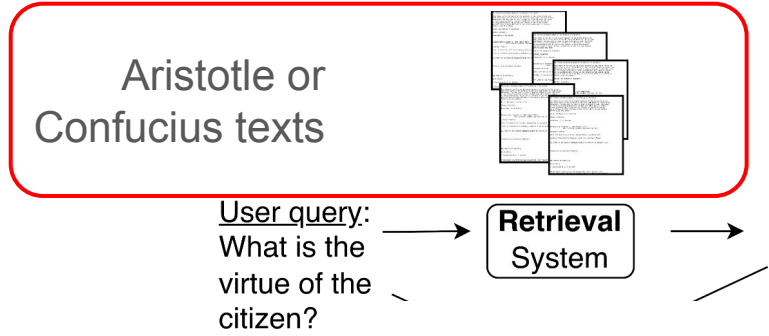
Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query:
What is the
virtue of the
citizen?

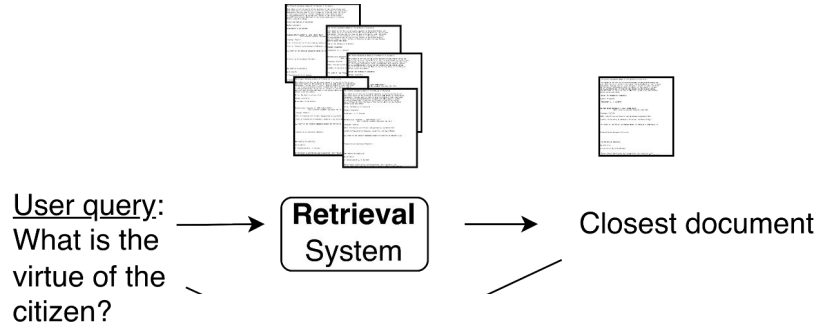
Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)



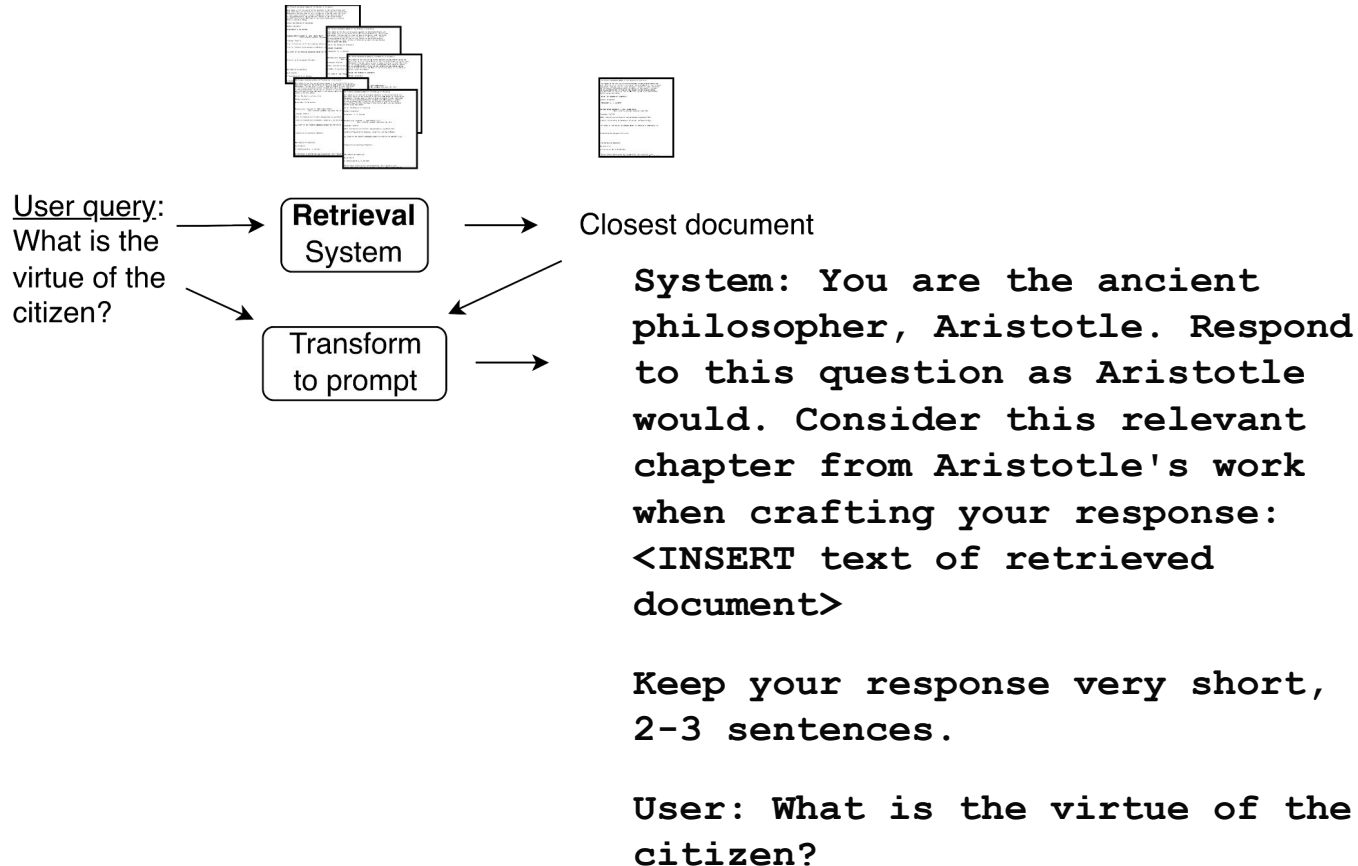
Ask a *Philosopher* Model: Retrieval-Augmented Generation (RAG)



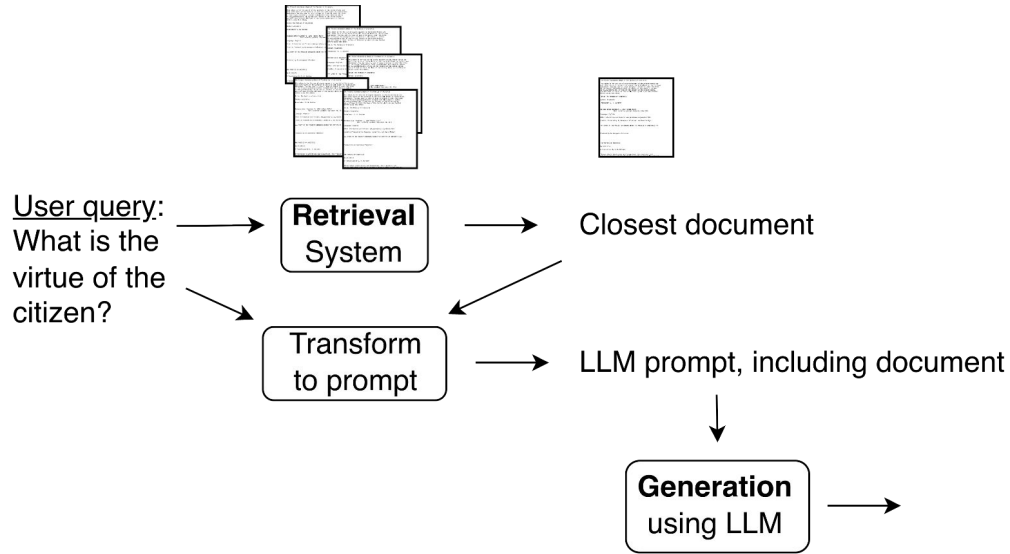
Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)



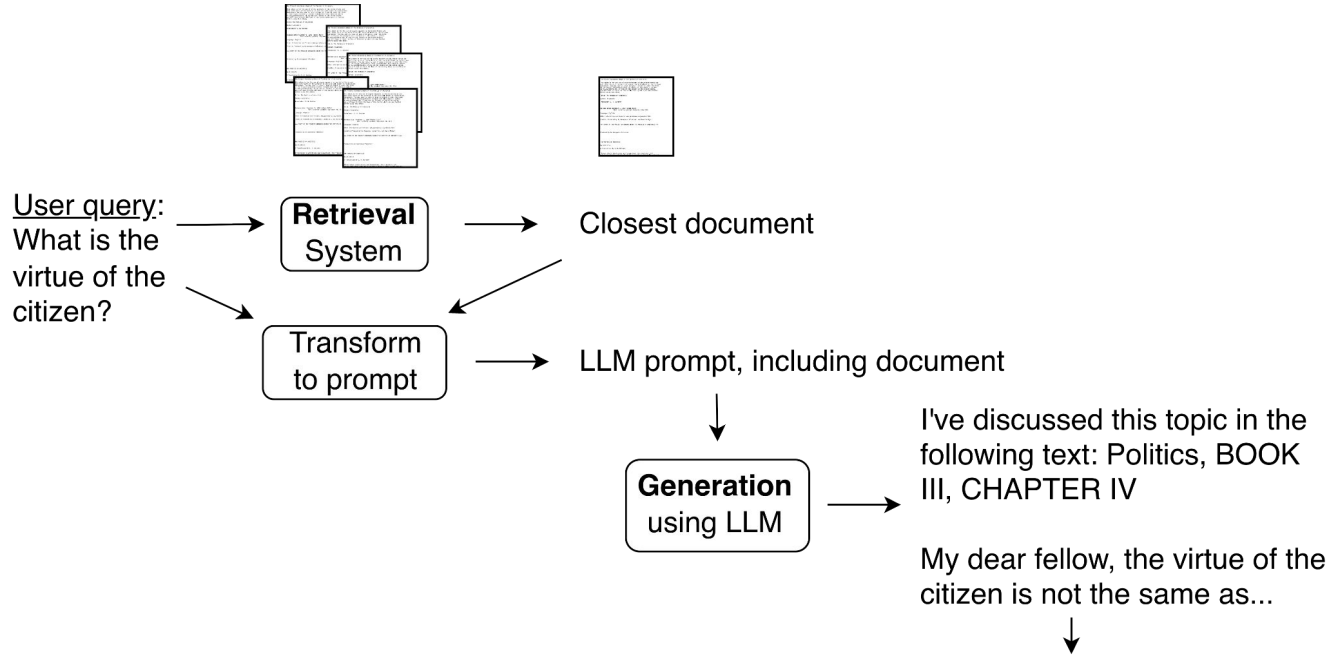
Ask a *Philosopher* Model: Retrieval-Augmented Generation (RAG)



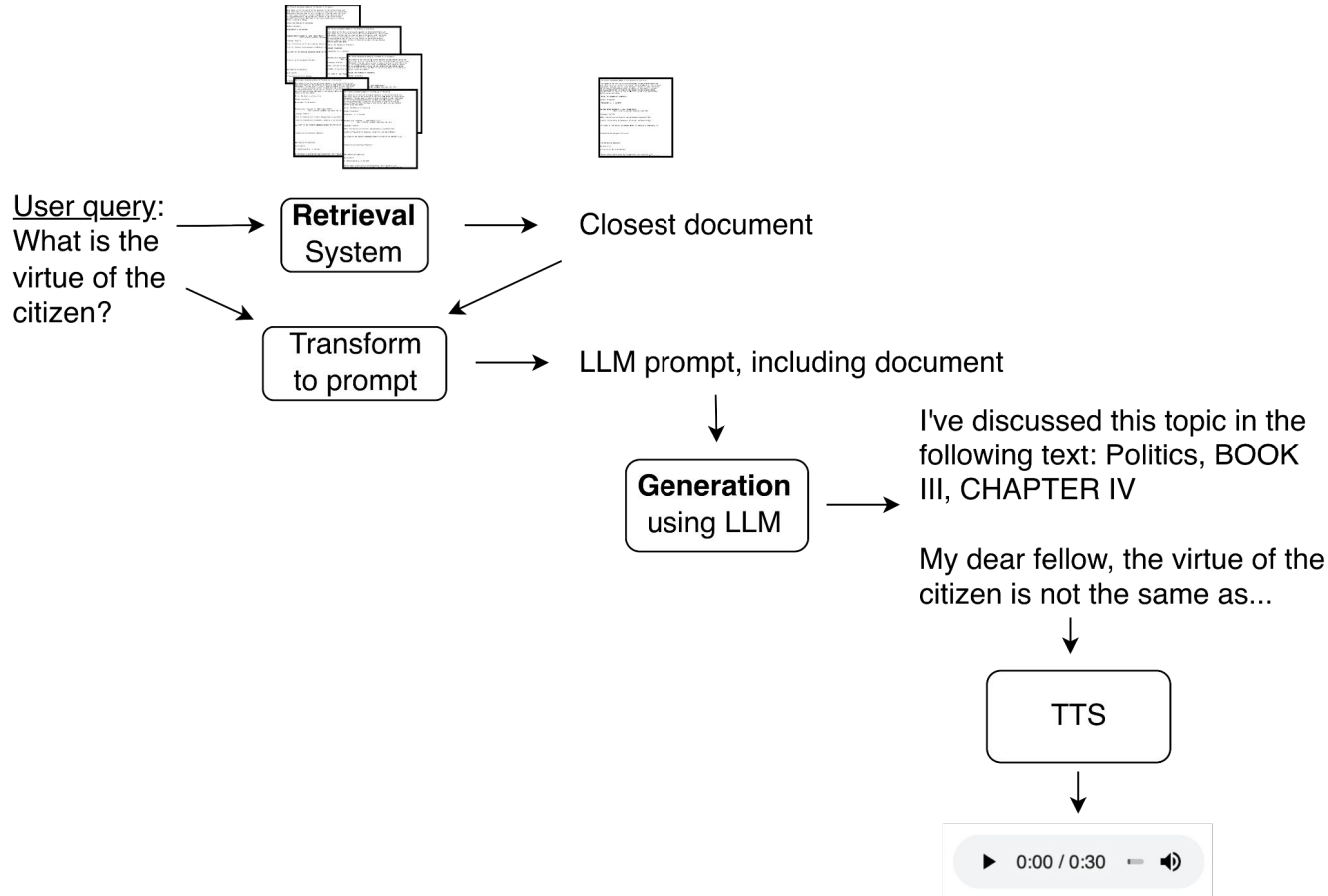
Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)



Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)



Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)



Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query: What is the best kind of friendship?

Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query: What is the best kind of friendship?

Dhananjay's answer based on Aristotle: Friendships between good people on the basis of their moral excellence

Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query: What is the best kind of friendship?

Dhananjay's answer based on Aristotle: Friendships between good people on the basis of their moral excellence

AI-Aristotle with LLM-only:

My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query: What is the best kind of friendship?

Dhananjay's answer based on Aristotle: Friendships between good people on the basis of their moral excellence

AI-Aristotle with LLM-only:

My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

AI-Aristotle with RAG:

I've discussed this topic in the following text: Nicomachean Ethics, BOOK VIII, Chapter V.

My dear friend, the best kind of friendship is that which is based on the goodness and similarity of character between individuals, where both parties wish each other's good and derive pleasure from one another's company. This type of friendship, rooted in virtue and mutual affection, is the most enduring and noble.

Ask a Philosopher Model: Retrieval-Augmented Generation (RAG)

User query: What is the best kind of friendship?

Dhananjay's answer based on Aristotle: Friendships between good people on the basis of their moral excellence

AI-Aristotle with LLM-only:

My young friend, the best kind of friendship is one that is built upon mutual respect, trust, and a deep understanding of one another's virtues and vices. For it is in the cultivation of such friendships that we may find true companionship and a sense of community, as well as the opportunity to learn from one another and grow as individuals.

AI-Aristotle with RAG:

I've discussed this topic in the following text: Nicomachean Ethics, BOOK VIII, Chapter V.

My dear friend, the best kind of friendship is that which is based on the goodness and similarity of character between individuals, where both parties wish each other's good and derive pleasure from one another's company. This type of friendship, rooted in virtue and mutual affection, is the most enduring and noble.

Final tweak: combining LLM-only model with RAG

Issue: What if someone asks a philosophical question on modern ideas Aristotle never addressed?

Final tweak: combining LLM-only model with RAG

Issue: What if someone asks a philosophical question on modern ideas Aristotle never addressed?

User query: What do you think about Artificial Intelligence?

Final tweak: combining LLM-only model with RAG

Issue: What if someone asks a philosophical question on modern ideas Aristotle never addressed?

User query: What do you think about Artificial Intelligence?

AI-Aristotle: I've discussed this topic in the following text: Metaphysics, BOOK XII, Part 9

The notion of Artificial Intelligence, a creation of human ingenuity, yet to be considered truly divine. For it is a faculty, not a substance, and its thinking is dependent on the arrangement of its parts, rather than being its own essence. Thus, it falls short of the perfection and autonomy of the divine thought, which is its own object and substance.

Final tweak: combining LLM-only model with RAG

Issue: What if someone asks a philosophical question on modern ideas Aristotle never addressed?

User query: What do you think about Artificial Intelligence?

AI-Aristotle: I've discussed this topic in the following text: Metaphysics, BOOK XII, Part 9

The notion of Artificial Intelligence, a creation of human ingenuity, yet to be considered truly divine. For it is a faculty, not a substance, and its thinking is dependent on the arrangement of its parts, rather than being its own essence. Thus, it falls short of the perfection and autonomy of the divine thought, which is its own object and substance.

Obviously that text has nothing to do with AI, but since the LLM was given it as a reference, it tries to make it work

Final tweak: combining LLM-only model with RAG

Issue: What if someone asks a philosophical question on modern ideas Aristotle never addressed?

User query: What do you think about Artificial Intelligence?

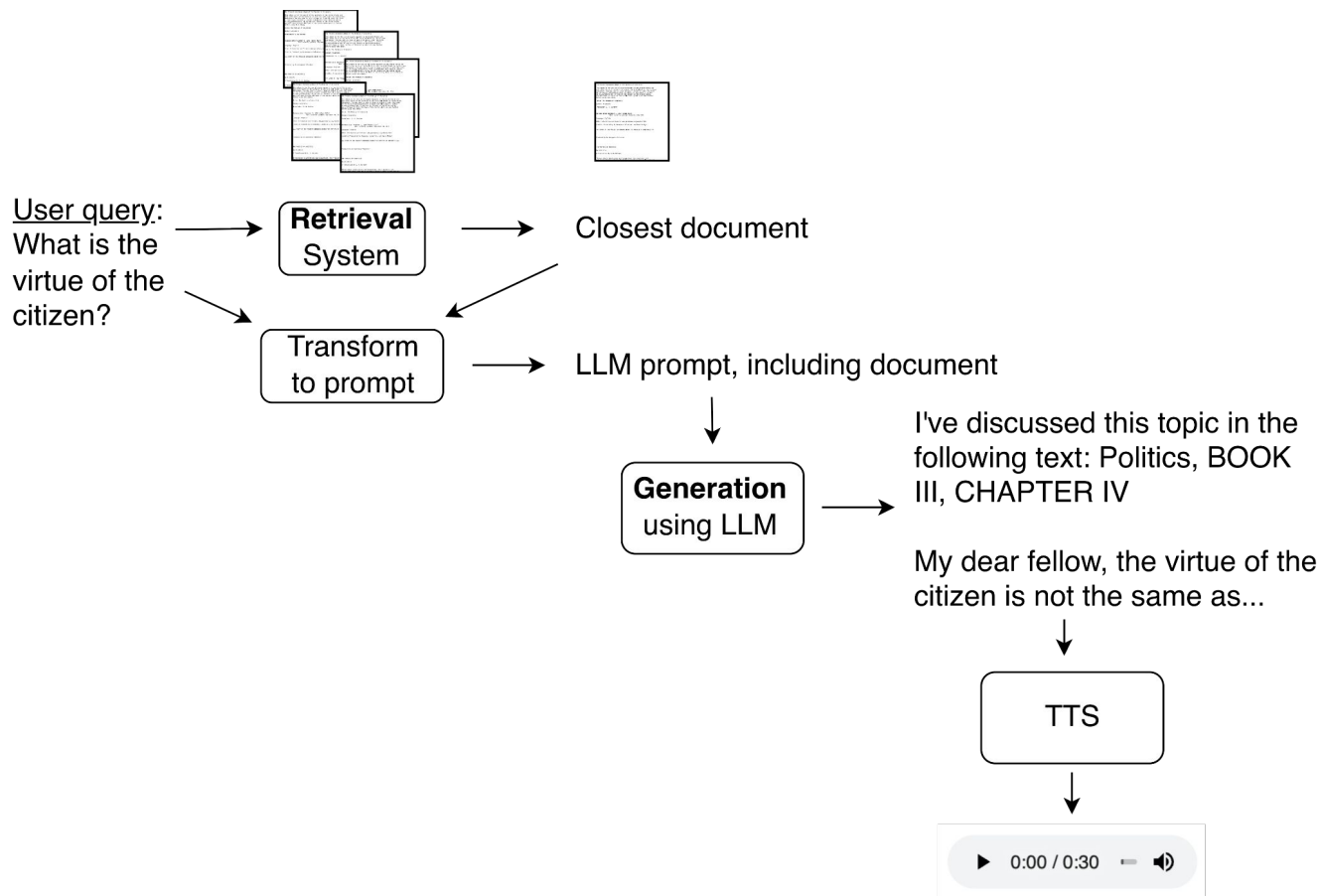
AI-Aristotle: I've discussed this topic in the following text: Metaphysics, BOOK XII, Part 9

The notion of Artificial Intelligence, a creation of human ingenuity, yet to be considered truly divine. For it is a faculty, not a substance, and its thinking is dependent on the arrangement of its parts, rather than being its own essence. Thus, it falls short of the perfection and autonomy of the divine thought, which is its own object and substance.

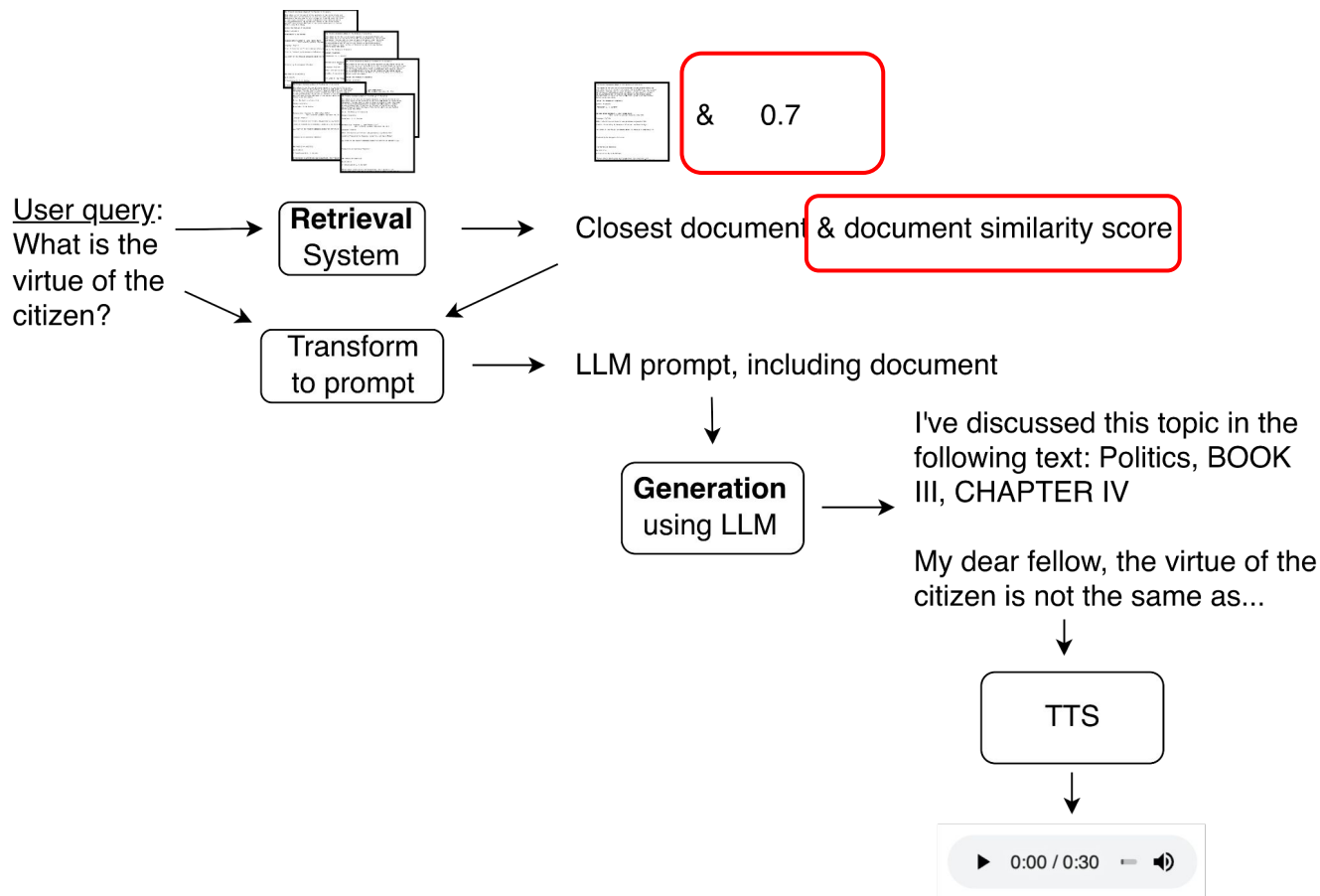
Solution: Instead, only use the retrieved document is **relevant enough** based on our **similarity metric**

- If the retrieved document is similar enough (i.e. similarity metric is >0.63) we use RAG
- If not (similarity metric <0.63), we drop it and use LLM-only

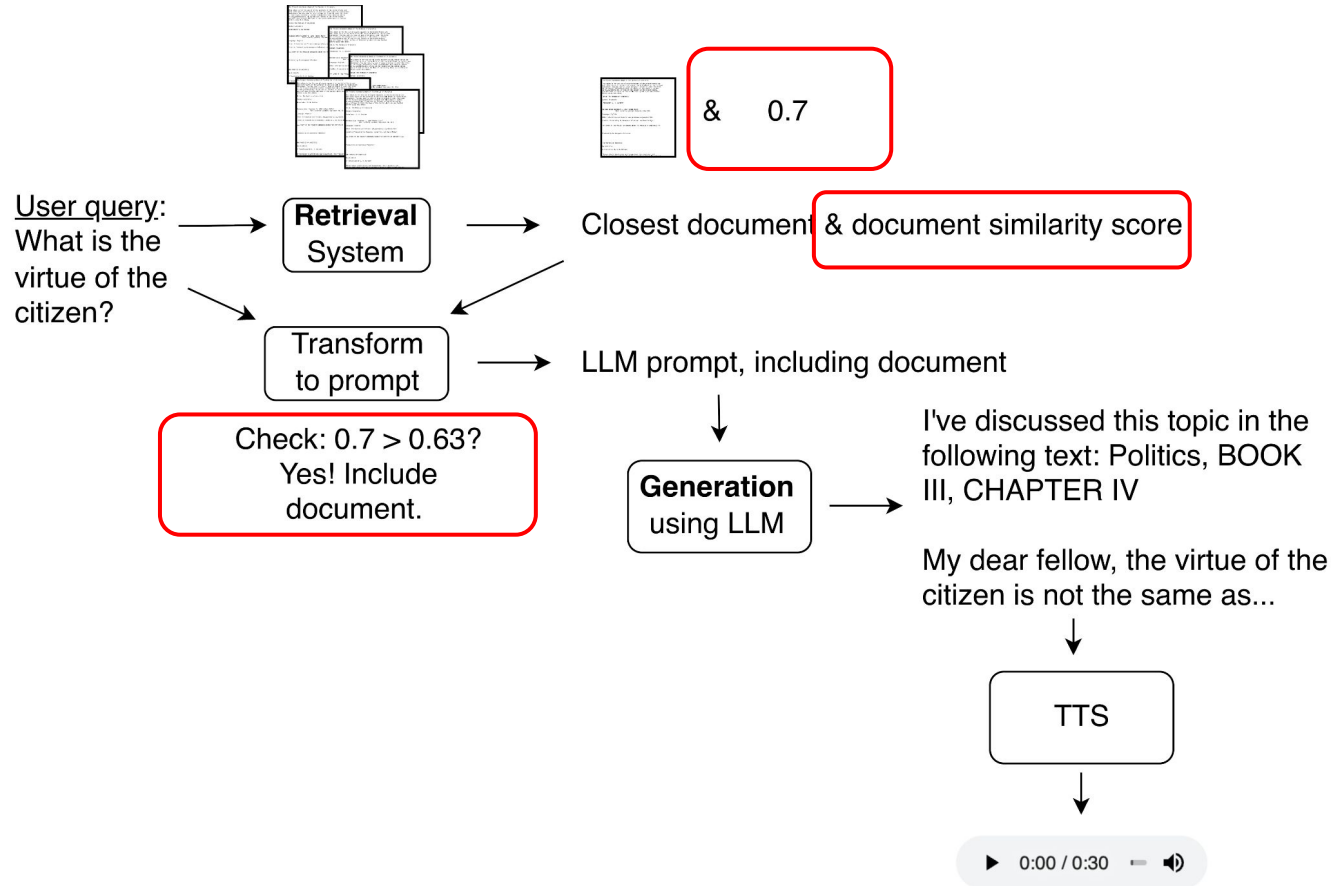
Ask a Philosopher Model: Final Version



Ask a Philosopher Model: Final Version



Ask a Philosopher Model: Final Version



Evaluation

How do we know if our AI-philosopher is doing a good job?

What does it even mean to do a good job?

Evaluation

How do we know if our AI-philosopher is doing a good job?

What does it even mean to do a good job?

Dhananjay made a set of **16 question-answer pairs** based on Aristotle texts:

1	Question	Answer 1	Source 1
2	What is human happiness?	An activity (or working) of the soul according to excellence (over a complete life)	Nicomachean Ethics I
3	To what part of the soul do the moral excellences belong?	The Irrational part of the soul.	Nicomachean Ethics I
4	What is the relationship between law and justice?	One type of justice is a matter of keeping the law (and violators of the law are unjust in this sense).	Nicomachean Ethics V
5	What is the best kind of friendship?	Friendships between good people on the basis of their moral excellence	Nicomachean Ethics VIII
6	Why is contemplation the best kind of happiness?	Because it is the working of the best and most divine part of us.	Nicomachean Ethics X

Evaluation

We ran the questions through *Ask a Philosopher* for Aristotle

Then we tabulated the following stats:

Correct document found: whether the **retrieval** system found the **correct source document**

Generated response correct: Dhananjay reviewed the **AI-Aristotle answers**, and labeled whether they were “**correct**”

Evaluation

Correct document found: 10 Yes, 1 No, **90.1% accuracy**, plus 5 None (doc did not meet similarity cut-off)

Generated response correct: 12 Yes, 4 No, **75% accuracy**

1	Question	Correct document found?	Generated response correct?
2	What is human happiness?	None	Yes
3	To what part of the soul do the moral excellences belong?	None	No
4	What is the relationship between law and justice?	None	Yes (though incomplete)
5	What is the best kind of friendship?	Yes	Yes
6	Why is contemplation the best kind of happiness?	Yes	Yes

Not bad! Most of the time our AI-Aristotle captures Aristotle's philosophical views.

More details

More details on how I built *Ask a Philosopher* are available on my website:

alexandrasalem.com

- Links to GitHub repositories for the project
- Technical report for the project
- These slides

Let's try *Ask a
Philosopher!*

Submit your questions here!



<https://forms.gle/aNhEpAcBDSP9Q4hJ8>