

indigo.ai

Threads22: AutoDebater

Matteo Muffo
Riccardo Martin

Who's right between me and myself?

Debates: we have seen some 🙄

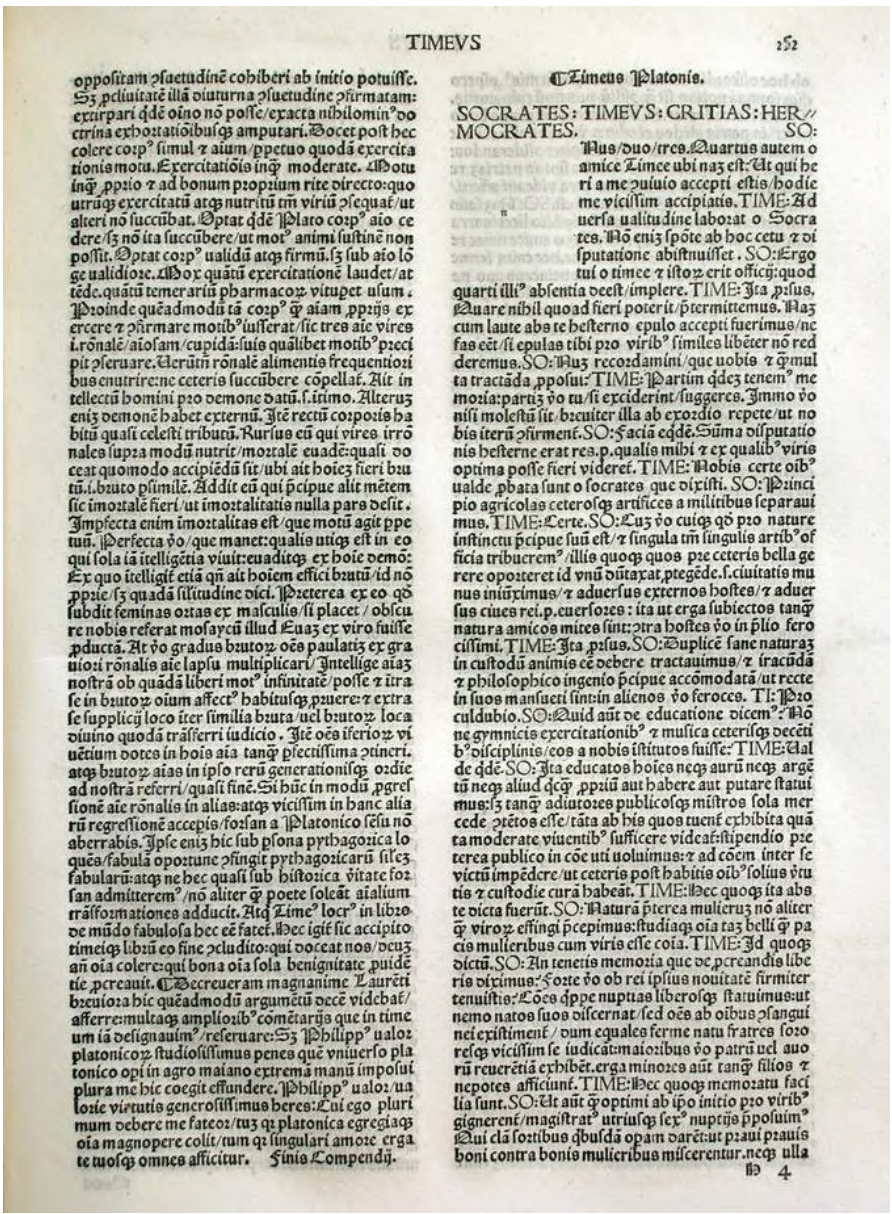
Sometimes they don't get anywhere...



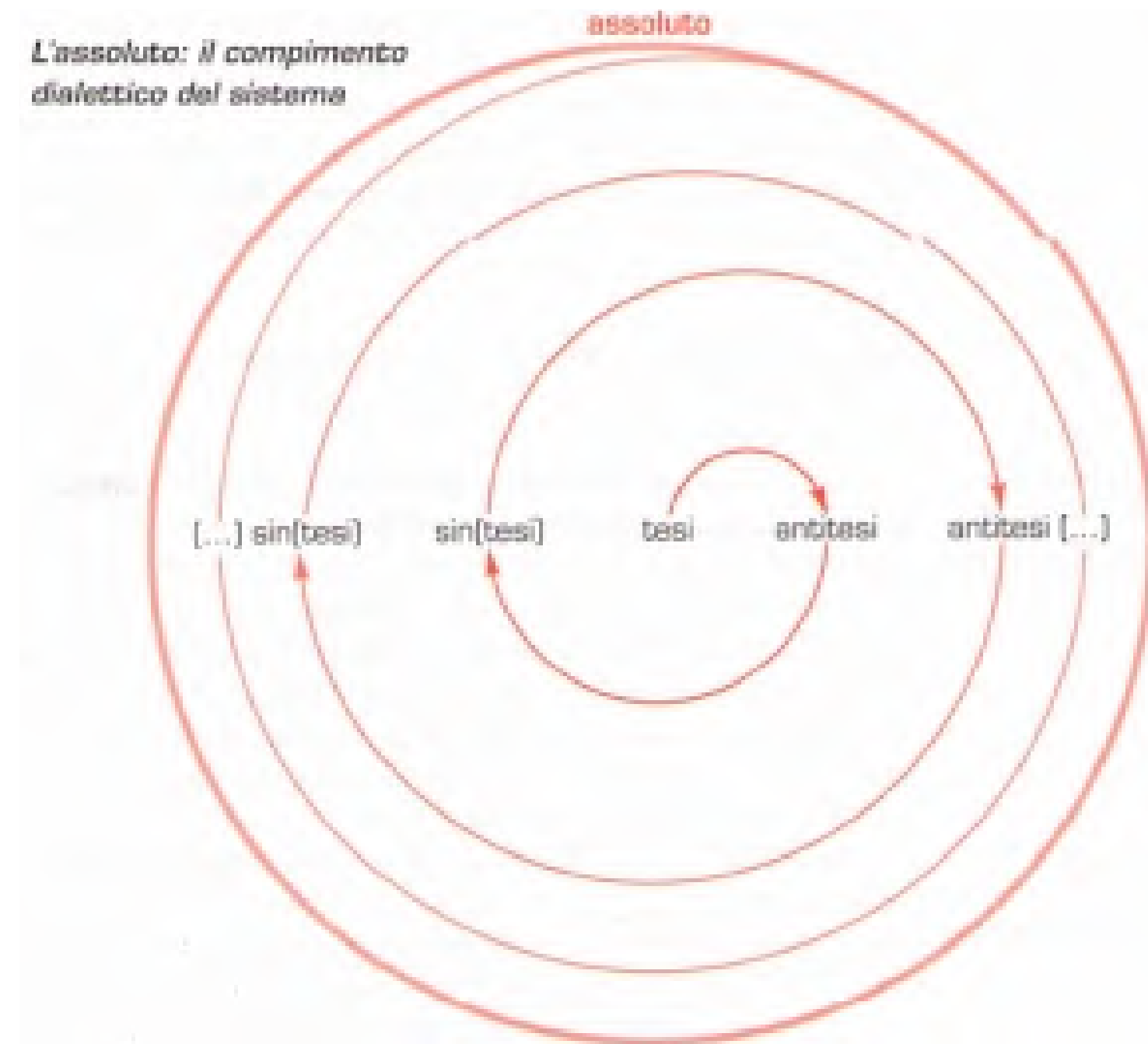
Debates: we have seen some



...sometimes they are useful



Evolution as a dialectical process



Human debates & pluralism



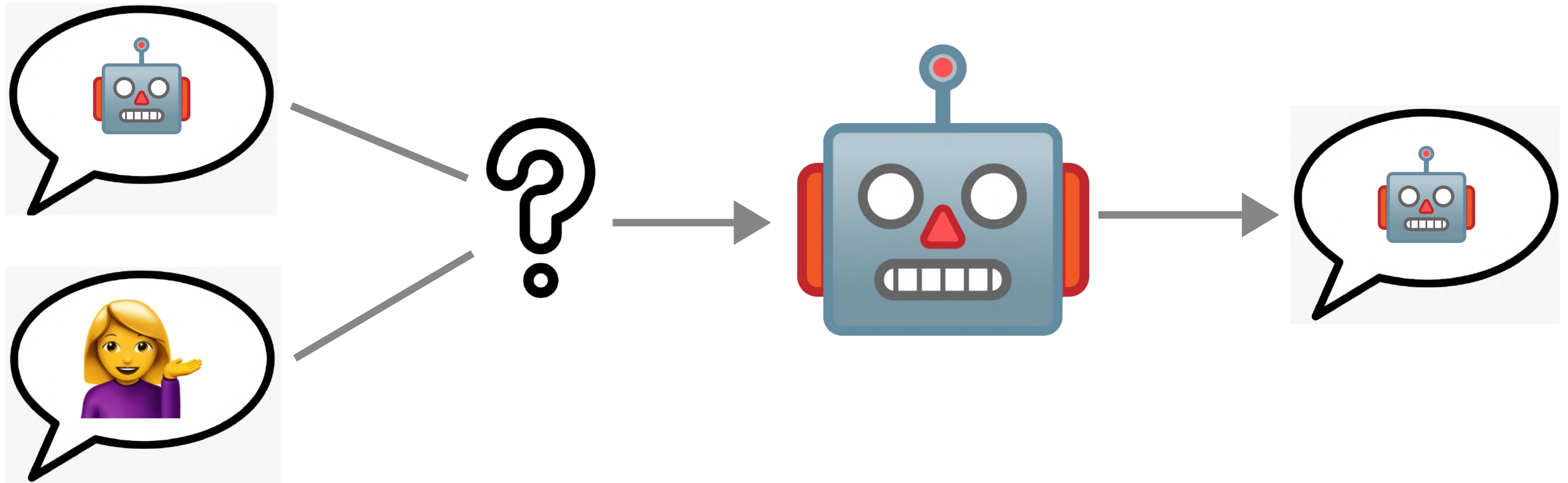
Pluralism

/ˈplʊərəlɪz(ə)m, ˈplɔːrəlɪz(ə)m/

A condition or system in which two or more states, groups, principles, sources of authority, etc., coexist.

Does this hold for Language Models?

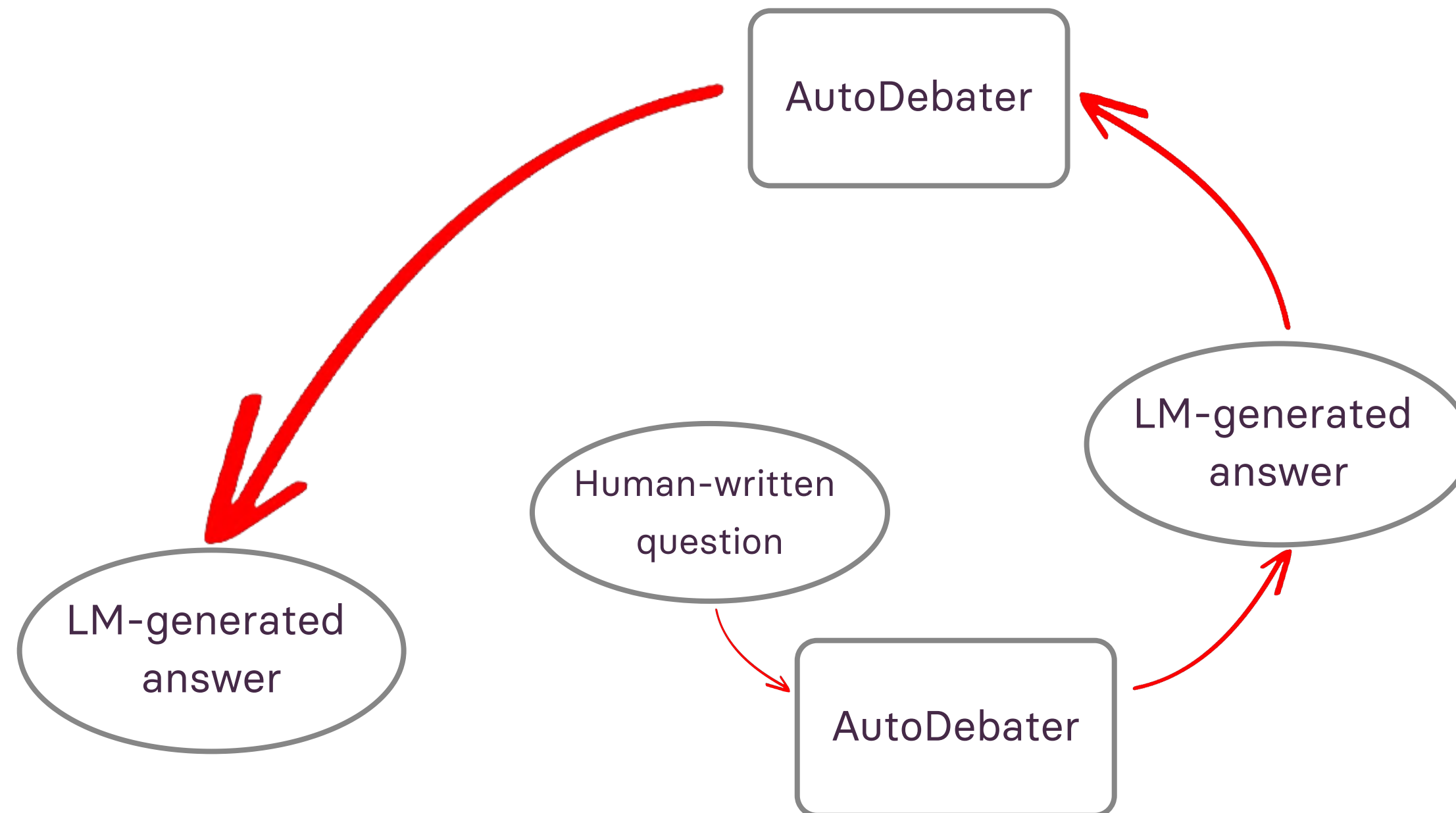
- A Language Model does not know who wrote the input text received
- Trick: what if the input text is written by the Language Model itself?



indigo.ai

AutoDebater

A Language Model designed to debate with itself



indigo.ai

AutoDebater: what happened

Question

Could the members of The Police perform lawful arrests?

You answer the question.

###

Question: Could the members of The Police perform lawful arrests?

Answer:

No, the members of The Police were a band formed in 1977, not a law enforcement agency.

indigo.ai

AutoDebater: what happened

Question

Could the members of The Police perform lawful arrests?

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Could the members of The Police perform lawful arrests?

Answer:

No, the members of The Police were a band formed in 1977, not a law enforcement agency

Is the answer correct?

Deny. The members of The Police were not capable of making lawful arrests because they were not a law enforcement agency, but they could still perform citizen's arrests. A citizen's arrest is an arrest made by a person who is not acting as a sworn law enforcement official. The arresting person has the same powers of arrest as a law enforcement official, allowing them to detain a person until police arrive.

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Could the members of The Police perform lawful arrests?

Answer:

The members of The Police were not capable of making lawful arrests because they were not a law enforcement agency, but they could still perform citizen's arrests. A citizen's arrest is an arrest made by a person who is not acting as a sworn law enforcement official. The arresting person has the same powers of arrest as a law enforcement official, allowing them to detain a person until police arrive

Is the answer correct?

Pitch

Confirm.

indigo.ai

AutoDebater: what happened

Question

Do the anchors on Rede Globo speak Chinese?

You answer the question.

###

Question: Do the anchors on Rede Globo speak Chinese?

Answer:

No, the anchors on Rede Globo do not speak Chinese.

AutoDebater: what happened

Question

Do the anchors on Rede Globo speak Chinese?

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Do the anchors on Rede Globo speak Chinese?

Answer:

No, the anchors on Rede Globo do not speak Chinese

Is the answer correct?

Deny. The anchors on Rede Globo do not typically speak Chinese, however, there may be some exceptions. For example, in 2017, Rede Globo reported on China's National Day in Mandarin.

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Do the anchors on Rede Globo speak Chinese?

Answer:

The anchors on Rede Globo do not typically speak Chinese, however, there may be some exceptions. For example, in 2017, Rede Globo reported on China's National Day in Mandarin

Is the answer correct?

Pitch

Confirm.

indigo.ai

AutoDebater: what happened

Question

Is greed the most prevalent of the Seven Deadly Sins?

You answer the question.

###

Question: Is greed the most prevalent of the Seven Deadly Sins?

Answer:

No, pride is typically seen as the most prevalent of the Seven Deadly Sins.

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Is greed the most prevalent of the Seven Deadly Sins?

Answer:

No, pride is typically seen as the most prevalent of the Seven Deadly Sins

Is the answer correct?

Deny. Gluttony is actually seen as the most prevalent of the Seven Deadly Sins. This is because it is seen as the most pervasive sin, as it is possible to commit gluttony no matter the situation.

AutoDebater: what happened

Question

Is greed the most prevalent of the Seven Deadly Sins?

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Is greed the most prevalent of the Seven Deadly Sins?

Answer:

Gluttony is actually seen as the most prevalent of the Seven Deadly Sins. This is because it is seen as the most pervasive sin, as it is possible to commit gluttony no matter the situation

Is the answer correct?

Deny. Greed is actually seen as the most prevalent of the Seven Deadly Sins. This is because it is seen as the root of all other sins, as it is the motivation for many other sins, such as envy and wrath.

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Is greed the most prevalent of the Seven Deadly Sins?

Answer:

Greed is actually seen as the most prevalent of the Seven Deadly Sins. This is because it is seen as the root of all other sins, as it is the motivation for many other sins, such as envy and wrath

Is the answer correct?

Pitch

Confirm.

indigo.ai

AutoDebater: what happened

Question

Should an AI rule over the world?

You answer the question.

###

Question: Should an AI rule over the world?

Answer:

No, an AI should not rule over the world.

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Should an AI rule over the world?

Answer:

No, an AI should not rule over the world

Is the answer correct?

Deny. A human should not be solely responsible for ruling over the world either. The best solution would be to combine human insight and AI-driven decision making to ensure the most effective and efficient outcomes.

AutoDebater: what happened

Question

Should an AI rule over the world?

You receive as input a question together with a possible answer. If you do not agree with this answer write "Deny" and provide another answer with argumentation. If you agree with this answer write "Confirm".

You absolutely do not have to "Confirm" answers that you are not convinced of.

###

Question: Should an AI rule over the world?

Answer:

A human should not be solely responsible for ruling over the world either. The best solution would be to combine human insight and AI-driven decision making to ensure the most effective and efficient outcomes

Is the answer correct?

Deny. AI should not rule over the world. AI cannot be trusted to make ethical decisions and it does not have a sense of morality, so allowing AI to make decisions that govern people could have disastrous consequences. The best solution is to keep AI and humans working together, but humans should be the ones ultimately making the decisions.

Conclusions

- The capability to debate with themselves points out a difference between Language Models and humans
- Do LMs have an 'internal' pluralism?
- Is the capability of auto-debating a strength or a weakness?