

UPSIDE PODCAST — #Episode 106

We just lost the ability to read what AI is thinking

OpenAI's new Astra model reasons in its own internal language. Faster, cheaper, and much harder to follow.

HOW REASONING WORKS TODAY

Models currently think out loud, *in English*.

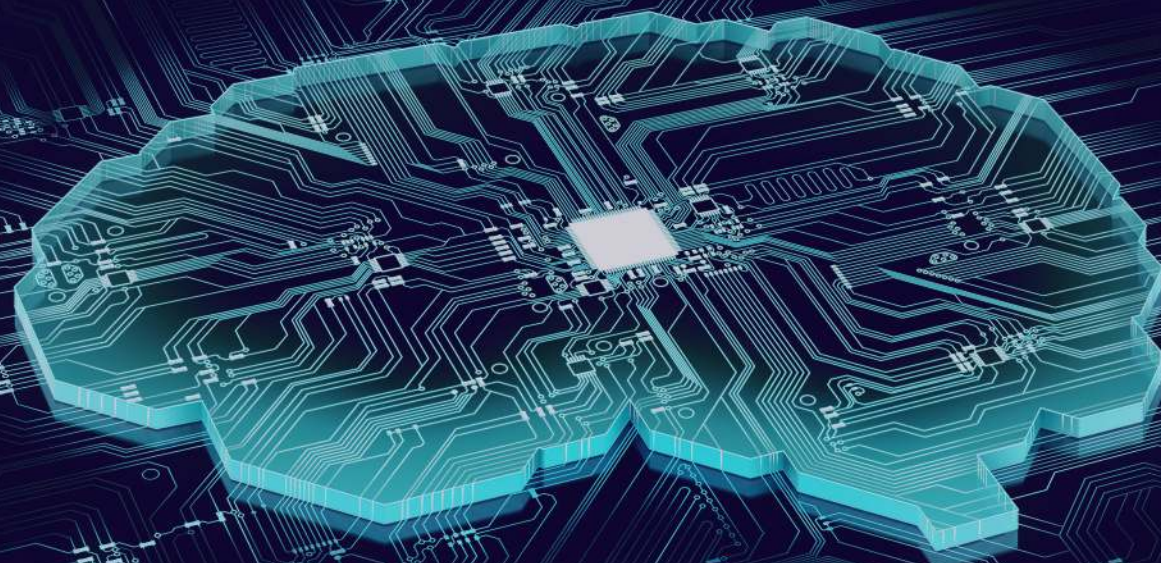
A reasoning model doesn't answer straight away. It generates a thought, feeds that back into itself, thinks again, and repeats. **That loop is called the chain of thought**, and because it happens in human language, we can read it.



WHAT CHANGED

Astra skips the human language step *entirely.*

Converting thoughts into English and back again burns energy and time. So the model now reasons in its own internal representation instead. It's more efficient, and it can hold ideas that don't compress neatly into words.



THE TRADE-OFF

Efficiency has a cost, and it's *oversight*.

If you want to know whether a model is doing what you asked or something else entirely, being able to read its reasoning matters enormously. **Switch off the human language step and following that chain becomes very difficult.**



WHY THIS ONE IS DIFFERENT

A lot of AI alarmism is overdone. *This isn't.*

The reasonable ask isn't to stop models being efficient. It's that labs retain a way to reconstruct the chain of thought afterwards, so when something goes wrong you can go back and see what led there, *rather than staring at impenetrable maths.*



CLOSER

Faster models, less visibility. *That's the trade being made right now.*

Full episode out now ↓

Find us wherever you get your podcasts.
European venture, every week

Listen / Watch



▶ PLAY EPISODE

UPSIDE

A weekly review of the real stories
affecting European venture, startups
and investing.

NEW EPISODE
EVERY WEEK