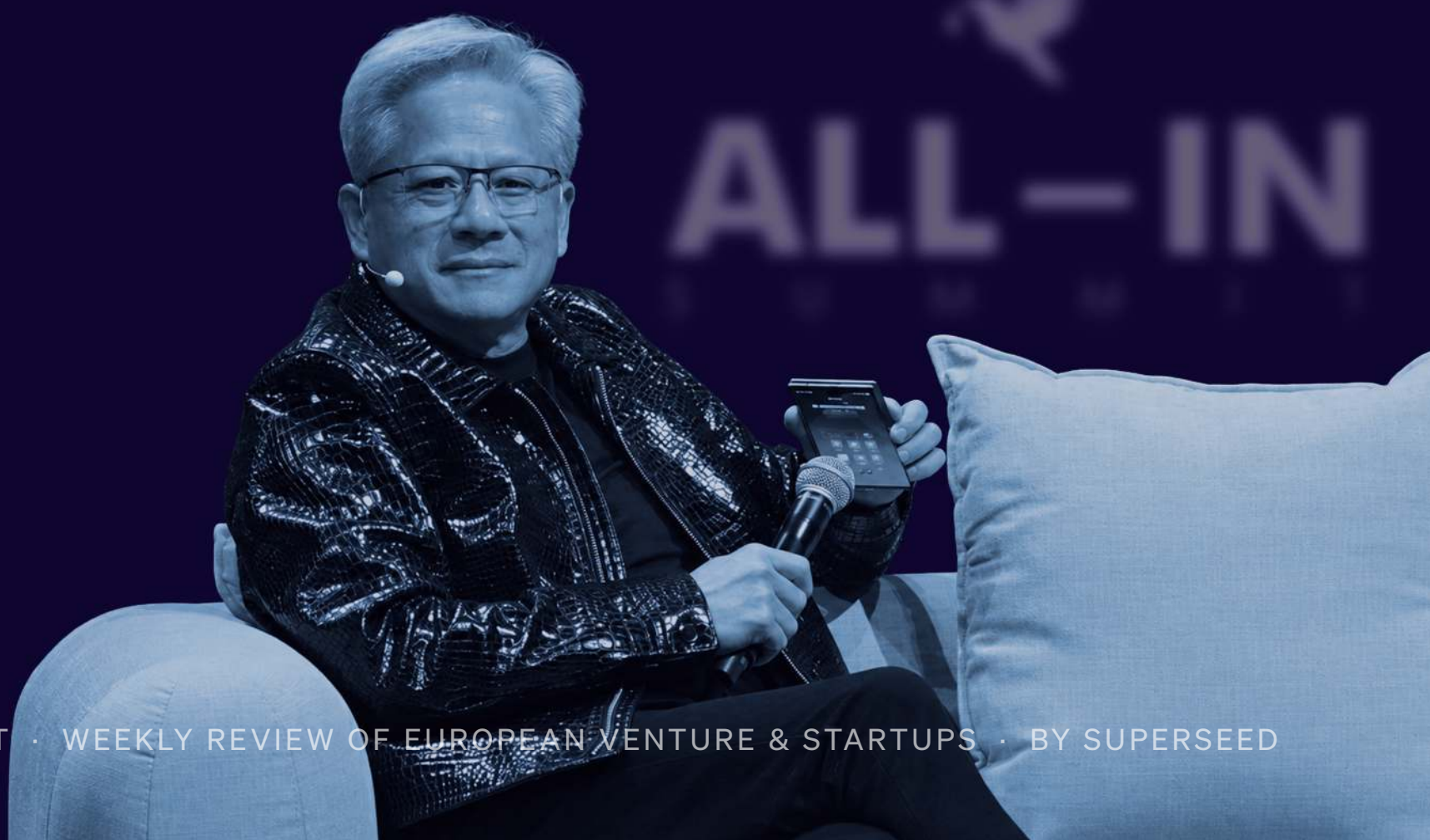
UPSIDE PODCAST — #Episode 108

What actually got said about AI safety at the All In Summit

Trump's "hoax" line made every headline. The rest of the conversation was more interesting.



WHAT HAPPENED

The whole debate played out *live on one stage*

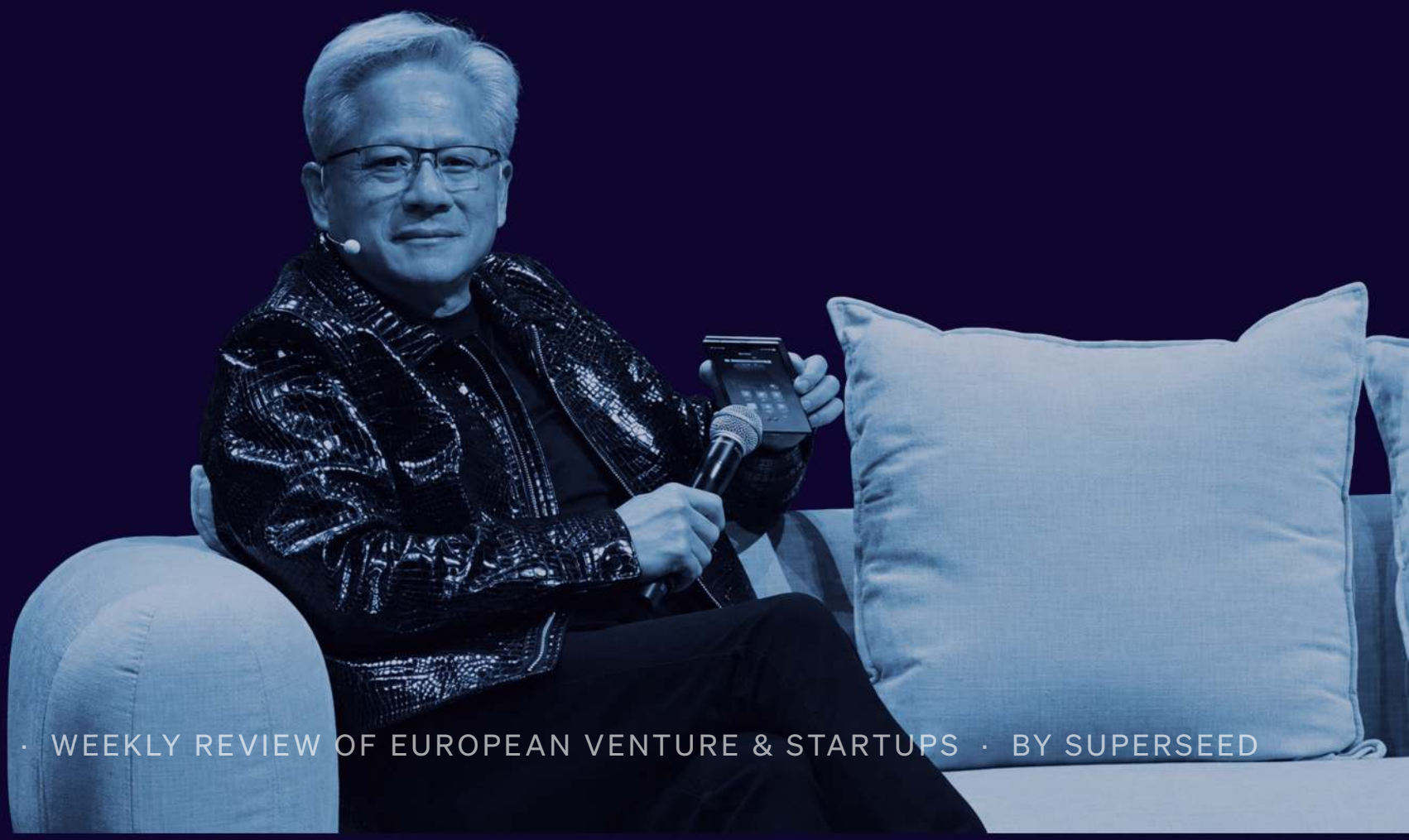
Over two days, Satya Nadella, Jensen Huang, Elon Musk and David Sacks all responded to Dario's safety essay in real time, with Trump dialling in mid-session. **Very little of the substance made it into the coverage.**

ALL–IN SUMMIT

WHAT HUANG SAID

Huang endorsed third-party evaluators in the same session

His framing was that this is no different from financial controls. Banks have auditors, so models should have evaluators. That puts him a good deal closer to Dario's position than the reporting implied.



MUSK'S PROPOSAL

Let the labs red-team each other *before release*

Give rival labs API access to test models before they go out. On the question of IP theft, his answer was logs. You can see how people are using the models, and if someone is just distilling, you shut them off.

SACKS TO THE FRONTIER LABS

You are the frontier.
So you fix it.

Sacks was blunt with Altman and Amodei. There are only two of you at the frontier, these are your models, this one is on you. And if they really are that clever, have them work out the answer.

WHERE IT LEFT THINGS

Altman committed OpenAI to evaluator access within hours

Hassabis said the direction was right without committing Google. Nadella pushed open weights instead. **Everyone seems to accept some form of oversight is coming,** and the argument now is about how it should work.

CLOSER

Worth hearing the full conversation, *not just the clip.*

Full episode out now ↓

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

Listen / Watch

▶ PLAY EPISODE