CIPA


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

# Cybersecurity oggi e domani

**Michele Colajanni**

Dipartimento di Informatica – Scienza e Ingegneria

michele.colajanni@unibo.it

# Diluvio di norme cyber

# Analisi degli attacchi del quinquennio 2020-2025

## **Colonial Pipeline** (2021)

**Riscatto**: 4,5M$ - **Danni**: >100 M$


## **Change Healthcare** (2024)

**Riscatto**: 18M$ - **Danni**: > 3 B$


## **CDK Global** (concessionarie auto)

**Riscatto**: 25M$ - **Danni**: > 1 B$

# Quale AI? Siamo in continua evoluzione

- Machine Learning
- Deep Learning
- AI generativa (LLM)
- AI agentica
- Self-evolving AI
- Autonomous Intelligence Systems
- Physical AI + Self-evolving AI
- …
- Artificial General Intelligence (AGI)

# Le frasi famose sono ancora valide?

**Emily Bender**

"AI is a stochastic parrot" (2021)

**Timnit Gebru**

"AI is all marketing" (2024)

# *Stochastic parrot?*

Art Made With Artificial Intelligence Wins at State Fair

Artist Jason Allen placed first in a Colorado contest, generating debate about A.I.'s role in art

Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad

Microsoft's AI Is Better Than Doctors at Diagnosing Disease

**The first trial of generative AI therapy shows it might help with depression**

The evidence-backed model delivered impressive results, but it doesn't validate the wave of AI therapy bots flooding the market.

Plots, love letters and remedies: The medieval secrets being revealed by AI

# Due informatici italiani – Due visioni diverse sull'AI

**Walter Quattrociocchi**

(Università Roma La Sapienza)

**Nello Cristianini**

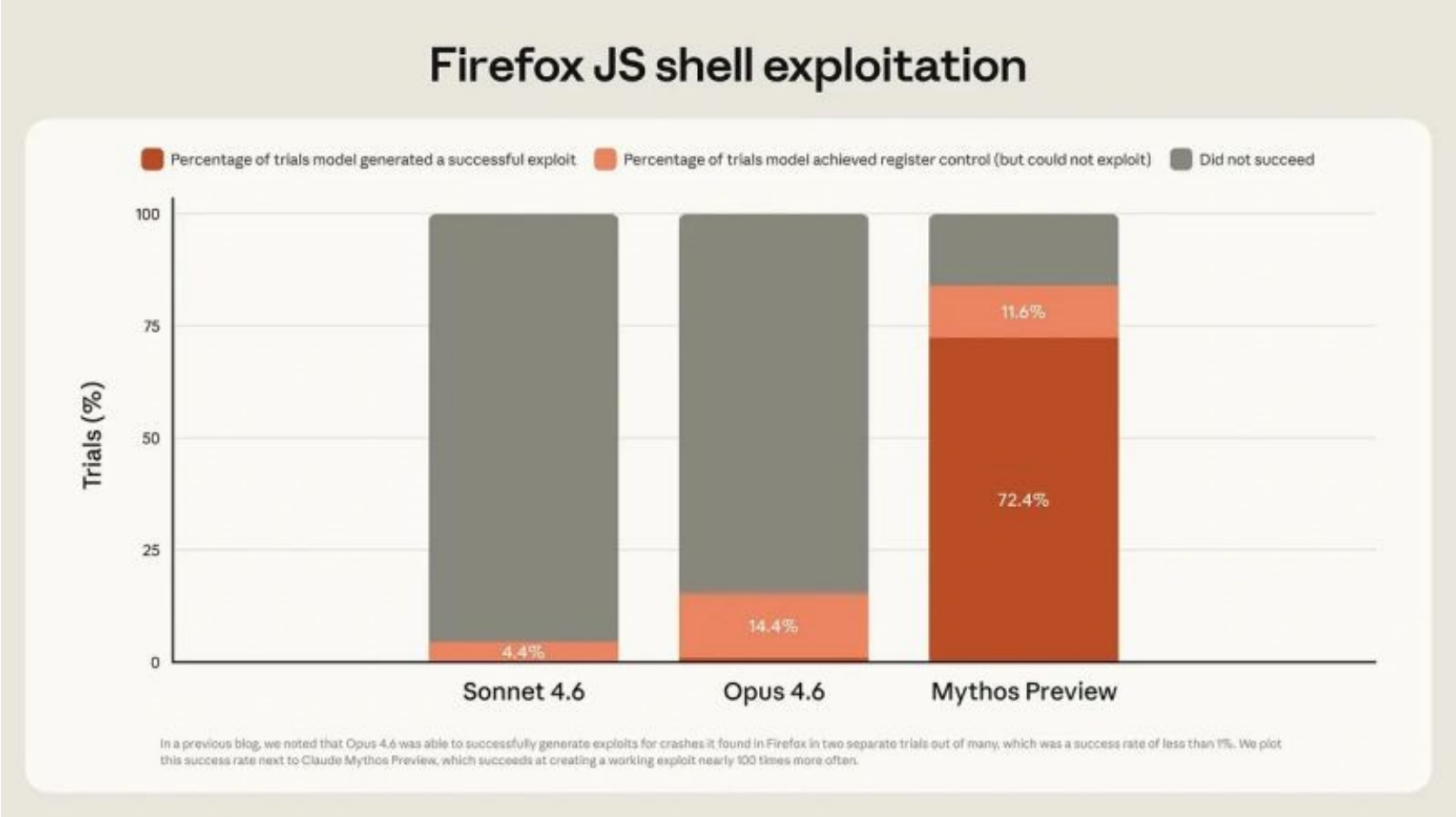
(Università di Bath, UK)

***Sono solo opinioni diverse o qualcosa sta cambiando?***

**Dagli storici modelli di Deep Learning (AlphaGo costituito da due reti con 200 e 400 neuroni → attuali modelli di Deep Learning con 10T di neuroni)**

# L’arrivo del cambio dirompente per la cybersecurity

# Mythos: quando la crescita diventa esponenziale



# Consentitemi dopo 25 anni un po’ di ottimismo

# Sta per cambiare tutto; anzi è già cambiato tutto

**DIFESA** - Un manipolo di esperti di sicurezza dotati di modelli di AI avanzati potrà valere quanto un intero team di *ethical bug bounty*

**ATTACCHI** - Purtroppo, anche gruppi criminali piccoli e relativamente competenti potranno acquisire capacità offensive prima riservate solo ad APT e apparati statali

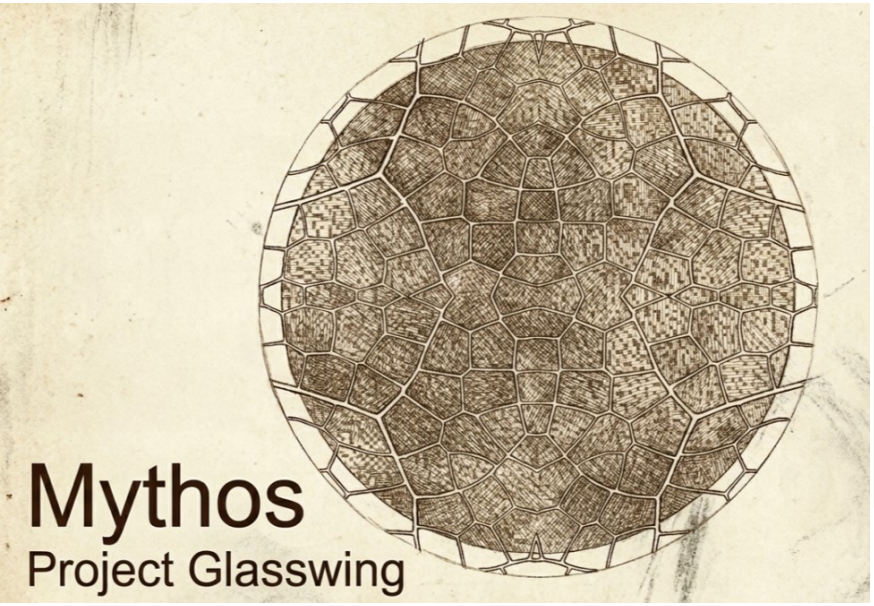
**AI Agents Enable Adaptive Computer Worms**

In our pursuit of new knowledge to enhance the security of artificial intelligence, we uncovered a cybersecurity threat with implications across society.

Andranno ripensati profondamente tutti i modelli di ***cyber risk management***

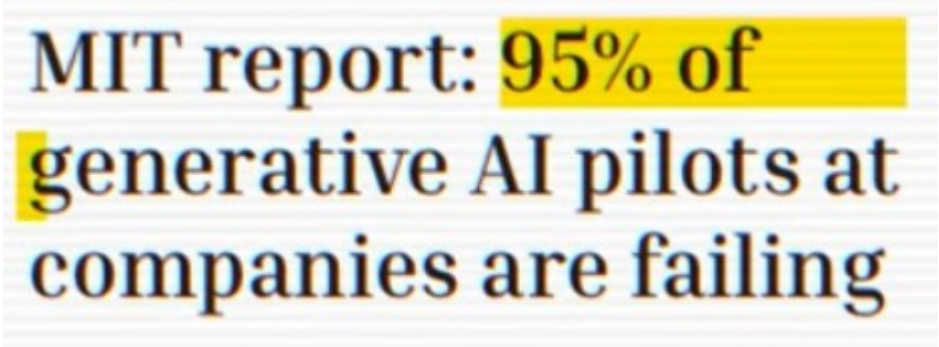
# E l'innovazione stimola la competizione

- Dopo Claude Mythos, appare ChatGPT 5.5 che rappresentano le attuali avanguardie dell'AI agentica, con prestazioni simili nei benchmark di cybersecurity e sviluppo software
- A differenza di Mythos, GPT-5.5-Cyber è rilasciato pubblicamente e ha costi di esecuzione un po' inferiori

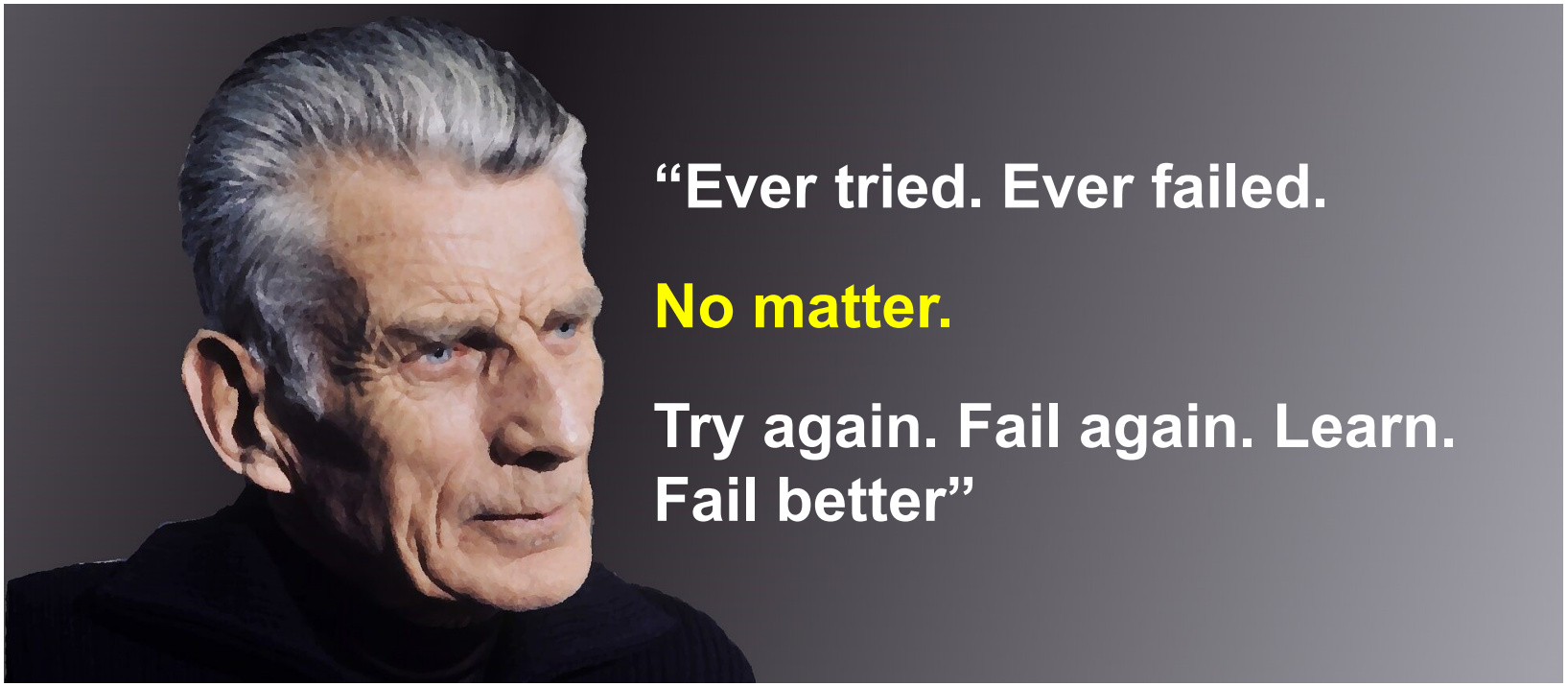



accenture BCG CROWDSTRIKE
Deloitte. Infosys Microsoft Security paloalto NETWORKS
pwc SentinelOne TrendAI WIZ

# La verità

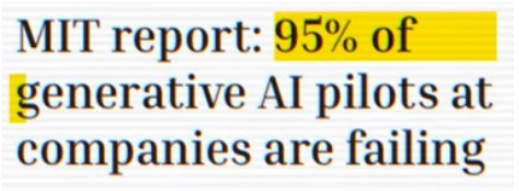

MIT report: 95% of
generative AI pilots at
companies are failing

# Approccio alla Samuel Beckett



# Motivazioni ben note

- **Progetto**: un'iniziativa temporanea intrapresa per creare un prototipo (PoC), un prodotto o un servizio unico
- **Programma**: in genere, un gruppo di progetti correlati gestiti in modo coordinato per raggiungere obiettivi strategici e benefici a lungo termine
- **Servizio**: un programma in grado di offrire un funzionamento continuo a terzi (interni/esterni) in modo fruibile




Le ragioni non sono tecnologiche, ma manageriali: investimenti motivati dall'hype del momento, ma due difetti noti:

- dati di partenza non strutturati e non continuativi
- mancanza di integrazione nei processi quotidiani

# Se vogliamo chiamarla in altro modo

# Con l'AI, la cybersecurity passa da attività artigianale a sistema industriale con nuovi processi e nuovi modelli

La cybersecurity smette di essere un'attività artigianale svolta da *mitici hacker esperti*, e tende a rientrare in un sistema di produzione industriale sempre più automatizzata e controllata.

➔ Una trasformazione *antropologica* ed *economica* analoga a quella subita dalla manifattura durante le rivoluzioni industriali:

- Cambiano i costi marginali, cambiano le barriere all'ingresso, cambia il rapporto tra capitale e lavoro, cambia il lavoro e quindi la formazione

# L’abisso della *trasformazione digitale*

**C’è il solito abisso tra la *digitalizzazione* e la *trasformazione digitale* che richiede modifiche nel management, nei processi e nella cultura aziendale**

**Il vero *digital divide* (così come la *cybersecurity* e l’AI) è sempre stato un problema cognitivo e manageriale**

MIT report: 95% of generative AI pilots at companies are failing

# Ottimismo con qualche recente perplessità

# Cloude Mythos → Oceanus(?)

Un recente leak del backend di Anthropic ha rivelato un cambiamento enorme nella loro strategia di prodotto. Mentre la comunità attende pazientemente segnali sul modello ristretto **Claude Mythos**, una stringa `claude-oceanus-v1-p` è trapelata insieme ai prezzi

**Prezzi Oceanus**: **$16/M token input - $80/M token output** (*il triplo di Claude Opus*)

Anthropic sta ignorando il mercato dei consumatori per riservare l'AI di alto livello ai Governi e alle imprese top Fortune 500

- Il framework Mythos non rilasciato non perché è "troppo pericoloso per la società", ma perché è un trampolino di lancio per costruire un ecosistema aziendale ultra-premium e blindato da cui estrarre il massimo profitto

Panico in Anthropic: l'intero programma Red Team è stato sospeso immediatamente in attesa di evoluzioni

# Residui

# In ogni caso, il 2026 è l'anno di Anthropic

**Feb**. - Anthropic si permette di proporre "limiti" al contratto illimitato col DoD, ma viene bandita come "fornitore inaffidabile"

**Mar**. – Data leak di Anthropic

**Apr**. – Produzione, ma non divulgazione di **Claude Mythos**; si preferisce il progetto *Glasswind* limitato a pochi grandi player

**15 Mag**. - Magnifica Humanitas, con la presenza di Christopher Olah

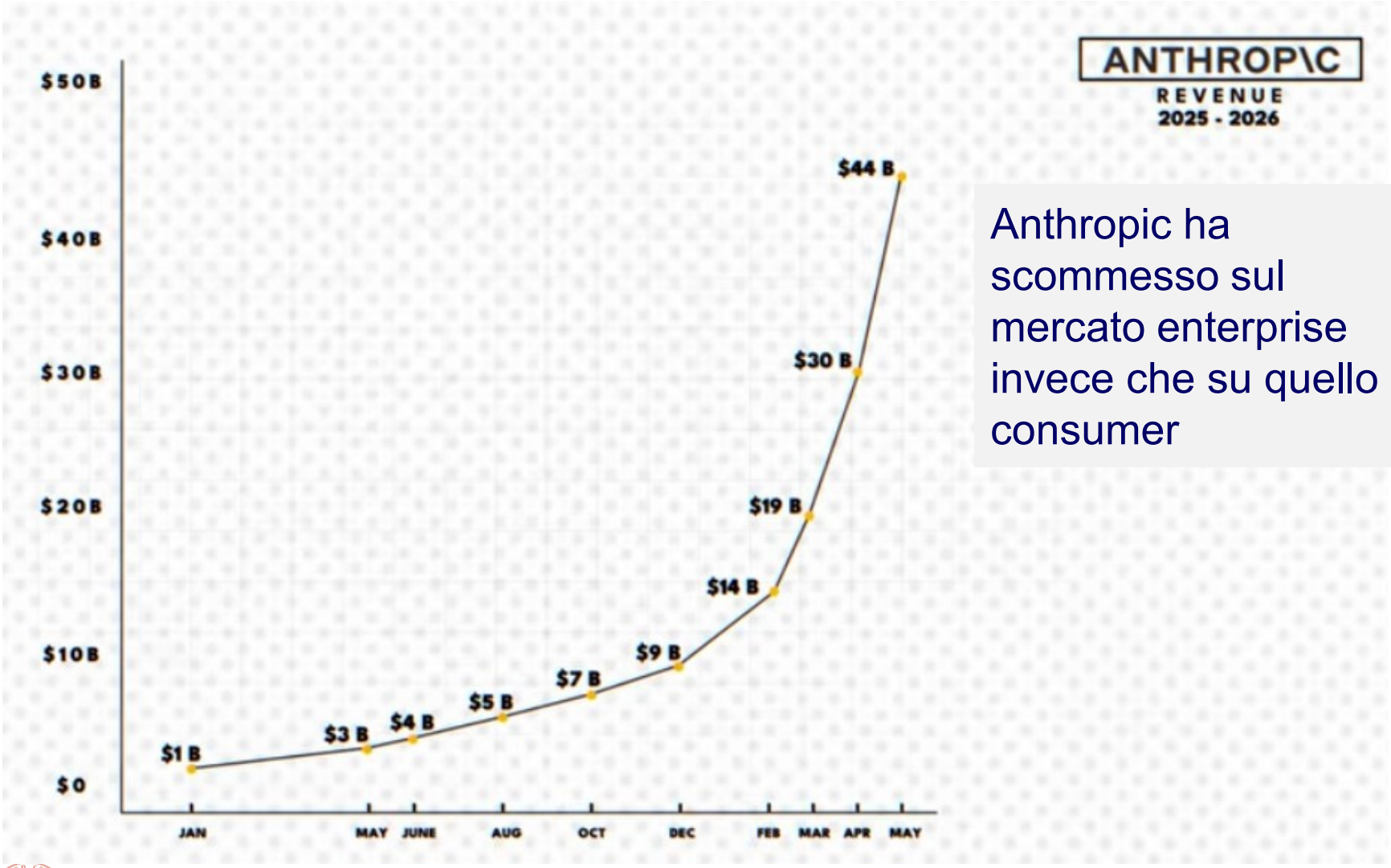
**31 Mag**. – *Initial Public Offering* per entrare in Borsa → 965B$

**5 Giu**. – L'AI potrebbe migliorarsi da sola (*Self-evolving AI*).

Si suggerisce una pausa prima che l'Intelligenza Artificiale inizi a migliorarsi in modi incomprensibili agli esseri umani

**Autunno 2026 (?)** – Ingresso in Borsa

# La società con più rapida crescita nella storia!




Anthropic ha scommesso sul mercato enterprise invece che su quello consumer

# Anthropic


Insane: 90-95% of Claude Code is now written by Claude Code.

I asked @mikeyk (Anthropic's CPO) what changes when so much of your code is written by AI?

His answer: The bottleneck moves from coding/design to:
1. Upstream: *aligning* the team around what to build.
2. Downstream: the PR *merge queue*, reviewing and merging all of the