Rad sa AI agentima

Šta donose, a šta oduzimaju

Vladimir Mijić

whoami

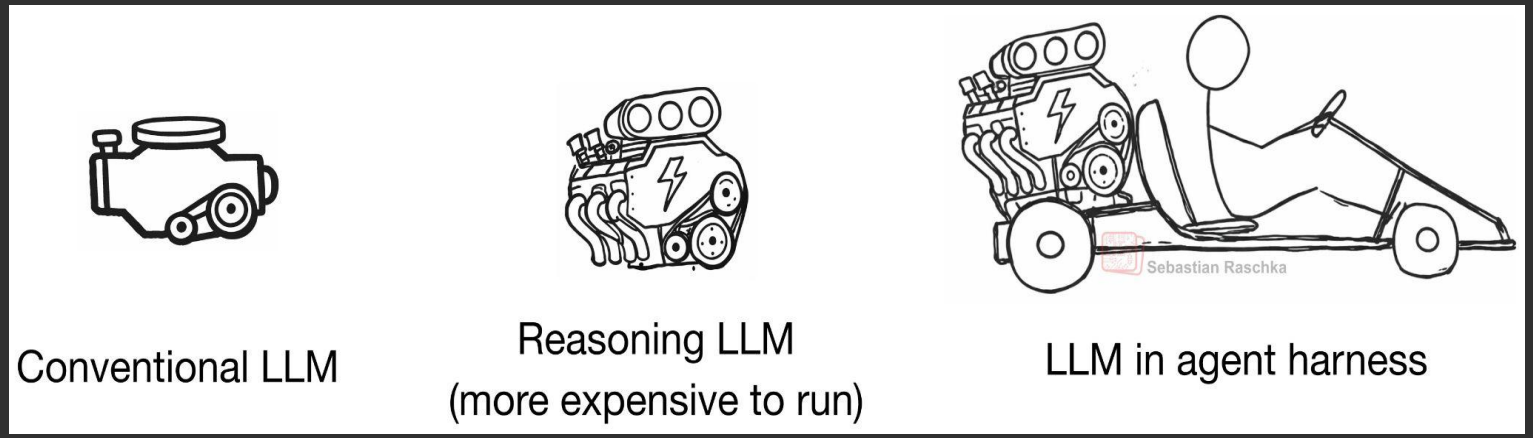
- Vladimir Mijić
- AI/ML inženjer @ LANACO
- BHFF alumni 

- Volonter @ CODE4SEE 

AI svakodnevica

- Od eksperimenta do navike - modeli su postali *commodity*.
- Interfejs se sporo mijenja - mijenja se ono što je u pozadini.
- Odgovori postaju akcije

Od LLM-a do agenta: promjena paradigme

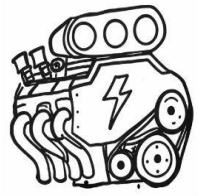


Sebastian Raschka - Components of A Coding Agent

Reasoning modeli: gdje daju najviše vrijednosti?

Najviše pomažu kada:

- problem zahtijeva više koraka
- cilj je jasno definisan
- rezultat možemo provjeriti



Reasoning LLM
(more expensive to run)

Nisu idealan izbor kada:

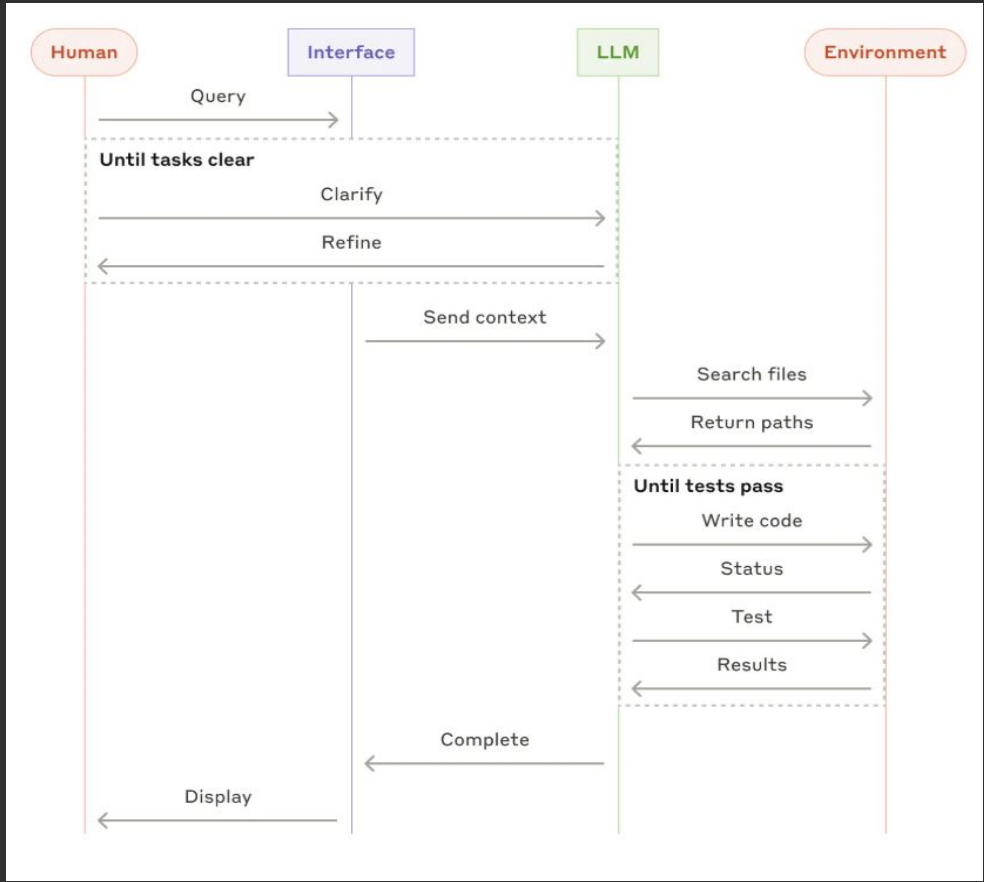
- zadatak je jednostavan
- važni su brzina i cijena
- ne postoji jasan kriterijum uspjeha

Sebastian Raschka - Components of A Coding Agent

Kako agenti rade?

Tipičan tok rada:

1. **Analiza zadatka**
2. **Prikupljanje konteksta**
3. **Korištenje dostupnih alata**
4. **Izvršavanje zadatka u okruženju**
5. **Provjera rezultata i iteracija**



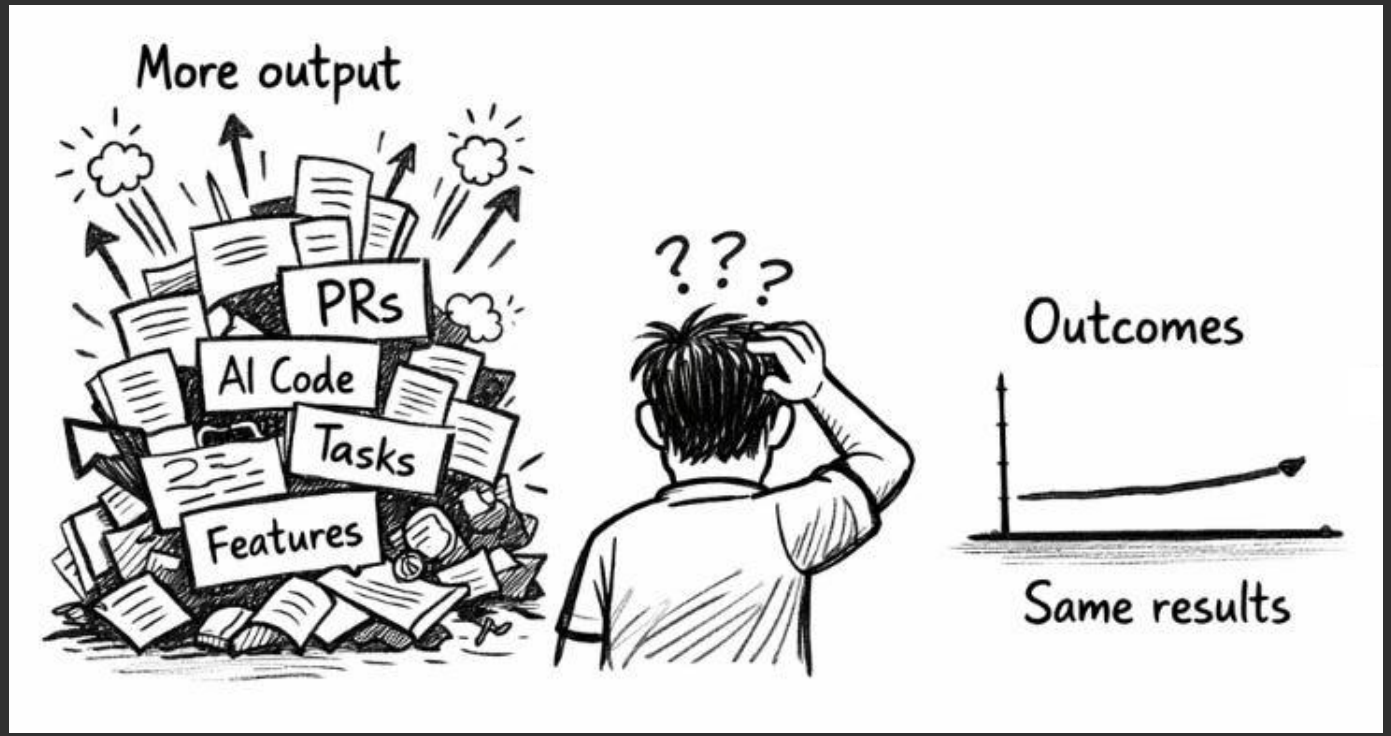
Anthropic - Building effective agents

Gdje agenti pomažu?

Najviše vrijednosti donose kod zadataka sa:
jasnim ciljem i **provjerljivim rezultatom**.

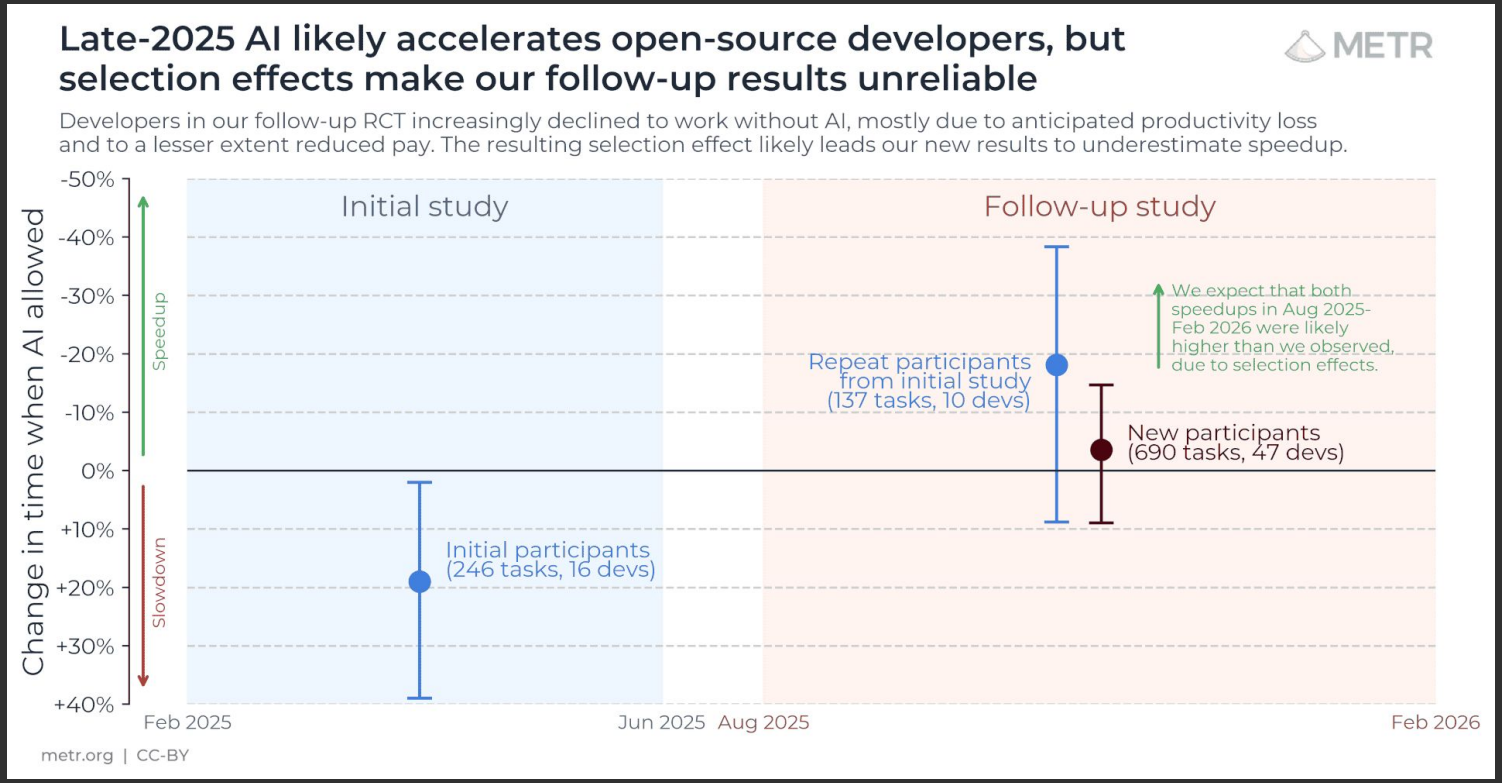
- razumijevanje postojećeg sistema
- implementacija i refaktorisanje manjeg opsega
- pisanje testova i pronalaženje propusta
- dokumentovanje izmjena

Da li smo zaista produktivniji?



The Engineering Tax - The Illusion of Speed

Šta pokazuju istraživanja?



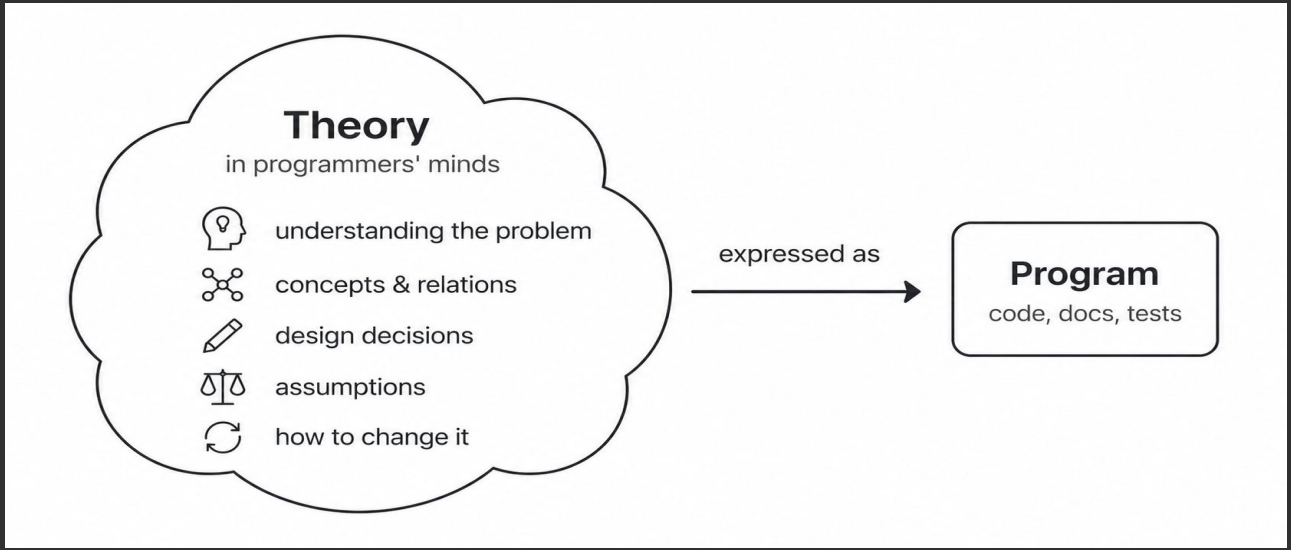
METER

Razvoj softvera nije samo pisanje koda

AI može ubrzati implementaciju.
Ali razvoj softvera zahtijeva da razumijemo:

- **šta sistem treba da uradi i zašto;**
- **koje su pretpostavke i kompromisi ugrađeni u rješenje;**
- **kako će se sistem dalje mijenjati i održavati.**

Razvoj softvera nije samo pisanje koda

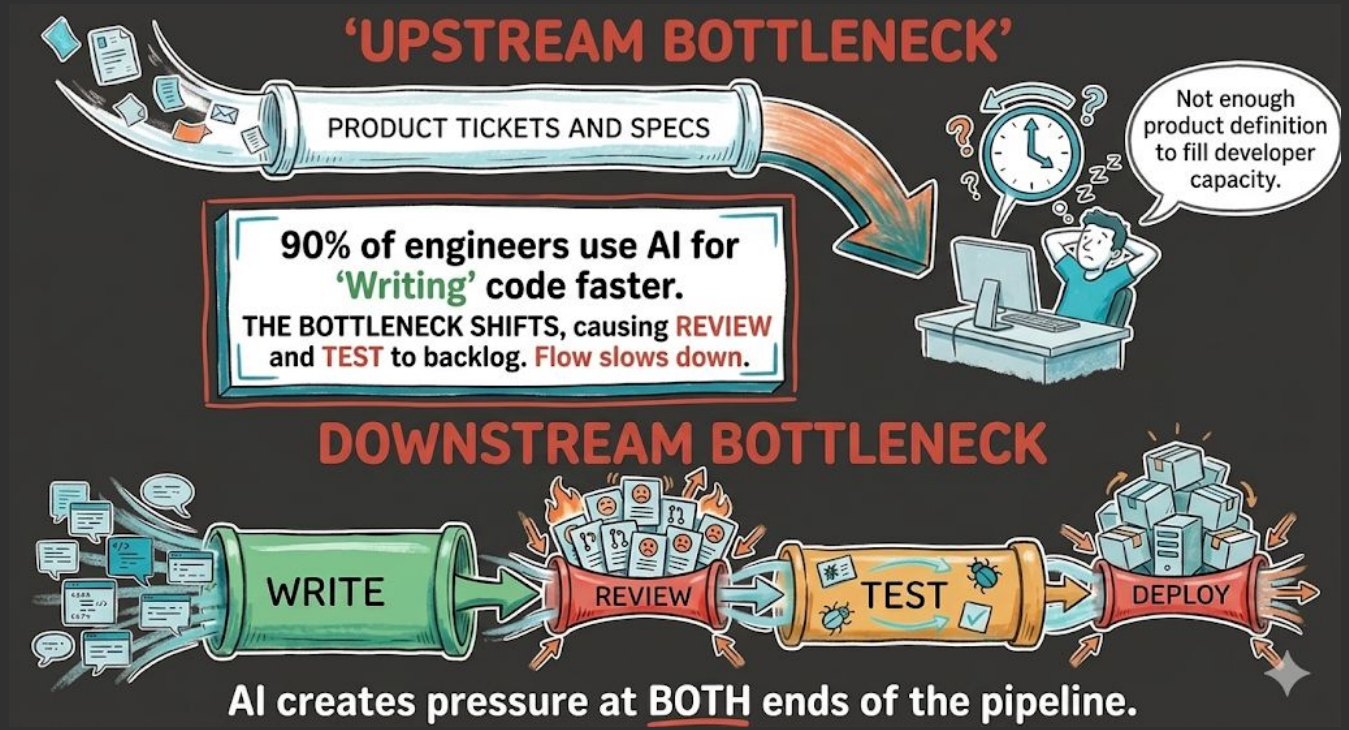


Ilustracija generisana koristeći GPT image 2

Peter Naur - Programming as Theory Building, 1985

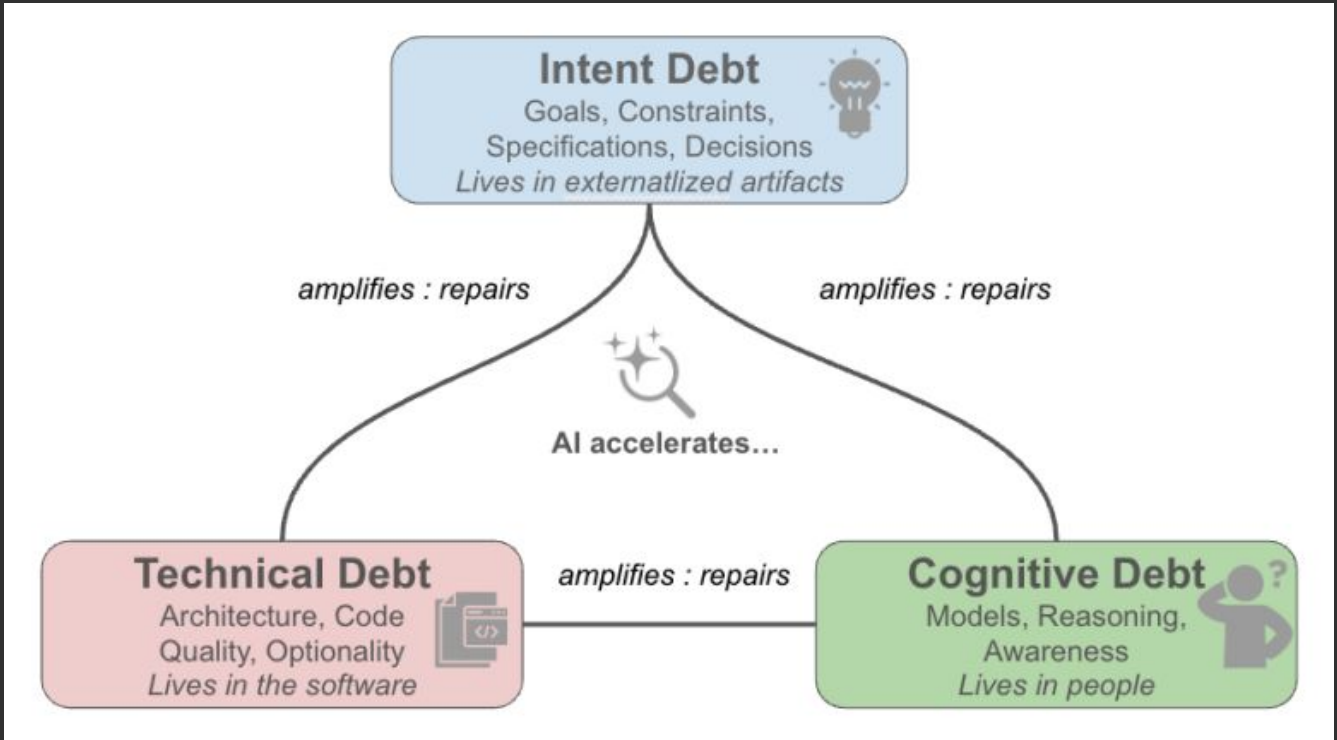
Usko grlo se pomjera

Više koda ≠ bolji sistem. AI možda može ubrzati pojedinca, ali ne i cijeli sistem!



Ilustracija generisana koristeći GPT image 2

Šta kada sistem raste brže od našeg razumijevanja?



Margaret-Anne Storey - From Technical Debt to Cognitive and Intent Debt, 2026

Kako raditi sa AI agentima?

AI kao alat, a ne prečica za učenje

AI može pomoći da naučimo ali može pomoći i da preskočimo učenje.

Dobra praksa	Loša praksa
Traženje objašnjenja	Preuzimanje gotovog rješenja
Traženje smjernica	Kopiranje koda bez razumijevanja
Poređenje različitih pristupa	Prihvatanje prvog ponuđenog rješenja
Pregled sopstvenog rješenja	Delegiranje bez analize problema

Judy Hanwen Shen and Alex Tamkin - How AI Impacts Skill Formation, 2026

Copilot, ne autopilot - 4D Framework

AI Fluency: Framework & Foundations

Zaključak

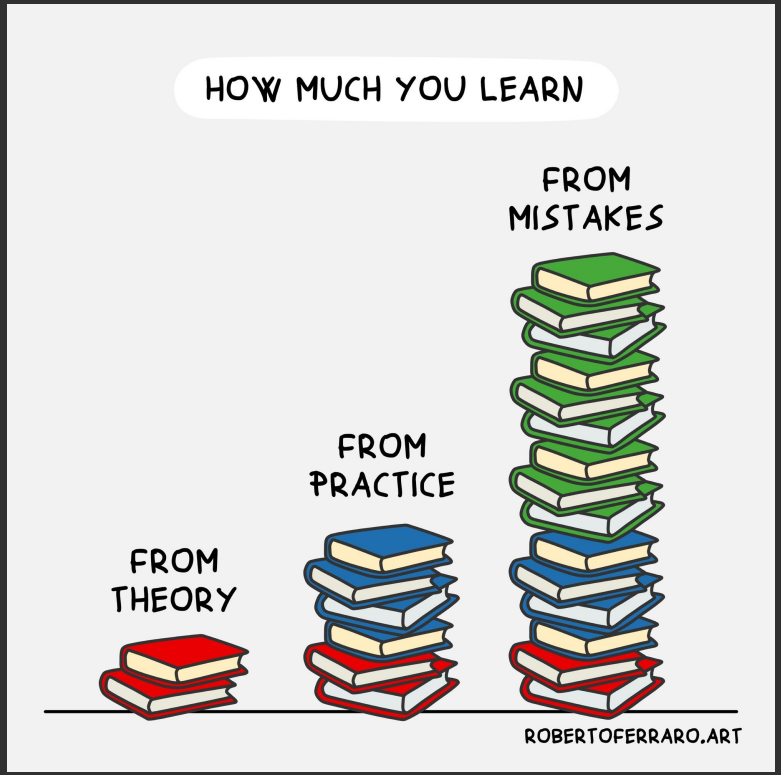
AI će nastaviti da smanjuje cijenu implementacije, ali razvoj softvera nikada nije bio samo proizvodnja koda.

Najvažnije je:

- razumijevanje problema i domena
- sposobnost procjene rješenja i njegovih posljedica
- očuvanje znanja unutar tima
- donošenje kvalitetnih odluka

Vrijednost inženjera se pomjera - sa brzine pisanja koda na kvalitet donesenih odluka.

Znanje se još uvijek isplati!



https://www.robertoferraro.net/illustrations

Hvala na pažnji!

Ostanimo u kontaktu

Vladimir Mijić

Backup slides…

Agent setup: Skills, Memory, Tools

https://agents.md/

https://developers.openai.com/codex/skills/

https://docs.anthropic.com/en/docs/claude-code/memory

https://docs.anthropic.com/en/docs/claude-code/mcp

https://cursor.com/docs/rules

Context engineering for Agents

Context Rot - https://research.trychroma.com/context-rot

How Long Contexts Fail - https://www.dbreunig.com/2025/06/22/how-contexts-fail-and-how-to-fix-them.html

Langchain Context Engineering
https://blog.langchain.com/context-engineering-for-agents/#:~:text=agent%20context%20engineering%20into%20four,is%20designed%20to%20support%20them

Effective context engineering for AI agents -
https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

Performance of models with large number of tools
https://www.catchmetrics.io/blog/a-brief-introduction-to-mcp-server-performance-optimization

The Berkley Functional Leaderboard
https://gorilla.cs.berkeley.edu/leaderboard.html

Copilot Researcher Agent
https://www.microsoft.com/en-us/microsoft-365/blog/2025/03/25/introducing-researcher-and-analyst-in-microsoft-365-copilot/

The Bitter lesson - https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

Data sources & reports

https://dora.dev/ai/roi/report/

https://newsletter.pragmaticengineer.com/p/ai-tooling-2026

https://newsletter.pragmaticengineer.com/p/the-impact-of-ai-on-software-engineers-2026

https://newsletter.pragmaticengineer.com/p/ai-impact-on-software-engineers-part-2

https://survey.stackoverflow.co/2025/ai/

https://arxiv.org/abs/2507.09089

https://getdx.com/blog/ai-assisted-engineering-q4-impact-report-2025/