PROGRAM VEČERA

Vibecoding Talks · pondělí 20. 4. 2026 · kino Dlabačov

17:30	*Otevřeno - neformální konverzace a hledání dobré židle*
18:00 - 18:30	**Patrick Zandl** Zahájení akce a Novinky v Claude Code
18:30 - 18:55	**Ondřej Tučný · Boldbrick** Strukturace dokumentace pro efektivní použití s Claude Code
18:55 - 19:10	*Přestávka 15 min*
19:10 - 19:40	**Jan Čurn · Apify** Pokaždé, když řekneš „CLI je lepší než MCP", zemře jeden 🦀 *Polemická rozprava o výhodách a nevýhodách CLI versus MCP*
19:40 - 20:10	**Martin König · Backlogica** Eliminace driftu systémů řízení práce za pomoci AI
20:10 - 22:30	**Patrick Zandl** Oficiální program končí *Volná debata se ze sálu postupně přesouvá do baru kina* 🍷

CLAUDE CODE 4.7

Duben 2026 · kritický přehled

Vibecoding Talks · 20. 4. 2026 · Patrick Zandl

vibecoding.cz | aivefirmach.cz

CO SE STALO ZA 6 TÝDNŮ

TECHNICKÉ NOVINKY

1. Opus 4.7 - xhigh effort, spotřeba tokenů
2. Scheduled tasks ≠ Loop
3. Computer Use (general availability)
4. Auto mode pro Max
5. Ultraplan / Ultrareview
6. Claude Design

BYZNYS KONTEXT

- Kauza AMD · issue #42796
- Regrese reasoning + adaptive thinking bug
- Prompt cache z 60 na 5 minut
- Pricing reforma pro enterprise
- BridgeBench re-ranking
- Konec paušální ekonomiky

Teze: novinky Opus 4.7 nejsou náhodné - jsou reakcí na tichou regresi a tlak na ekonomiku.

1

OPUS 4.7

Víc přemýšlení, víc tokenů, víc peněz.

OPUS 4.7 vs 4.6 (a Sonnet 4.6)

EFFORT LEVELS

+2

xhigh + max

Nové stupně reasoning budgetu nad rámec high. Určeno pro deep reasoning.

SPOTŘEBA TOKENŮ

~2×

na turn vs. 4.6

Max effort alokuje mnohonásobně více thinking tokenů. Default se nezměnil.

SONNET 4.6

=

beze změny

Opus 4.7 není náhrada Sonnetu. Sonnet zůstává default pro rychlé iterace.

Kritická poznámka: ‚xhigh" a ‚max" nemají v release notes jasný token budget. Billing se řídí skutečnou spotřebou.

SCHEDULED TASKS ≠ LOOP

Dvě věci, které vypadají podobně a dělají zcela něco jiného.

SCHEDULED TASKS

Naplánovaný úkol v čase - cron pro Claude.

- Běží jednou / opakovaně podle intervalu
- Každé spuštění = nový kontext, nová session
- Vhodné: monitoring, report, daily briefing
- Nevhodné: iterativní hledání řešení problému

LOOP / AGENTIC LOOP

Claude iteruje nad problémem, dokud neskončí.

- Sdílený kontext přes mnoho turns
- Běží dokud není cíl splněn nebo vyčerpán budget
- Vhodné: refactor, debug, code generation
- Vhodné: generování reportu každé ráno

Pozor na záměnu: scheduled task pro ‚ráno mi shrň issues" ≠ nekonečný debug loop. Druhé vám vyžene účet.

KAUZA AMD · ISSUE #42796

Stella Laurenzo (Senior Director AI Compiler, AMD), 2. 4. 2026 - 6 852 sessions, 234 760 tool calls, 17 871 thinking bloků.

-67 až -73 %

DÉLKA REASONINGU

Medián ~2 200 → ~600 znaků

6,6 → 2,0

READS / EDIT

Edituje ‚naslepo"

0 → 173

STOP-HOOK VIOLATIONS

Za 17 dní po 8. 3.

3×

SEBEKONTRADIKCE

Halucinace: git SHA, apt balíčky, API

> *„Claude has regressed to the point it cannot be trusted to perform complex engineering."*
>
> Stella Laurenzo · AMD tým přešel k jinému poskytovateli (NDA)

Slabiny studie: hypotéza o GPU-load-sensitive allocation je interpretace, Anthropic odmítl. 80× nárůst API requestů je částečně artefakt škálování z 1-3 na 5-10 souběžných agentů.

REAKCE ANTHROPIC · ZERO-REASONING BUG

1 Fáze 1 · GitHub

Boris Cherny (vedoucí Claude Code): ‚defaultní effort se 3. 3. posunul z high na medium, použijte /effort high."

Laurenzo: ‚už dávno effort=max, nepomáhá."

2 Fáze 2 · Hacker News

Přiznání: adaptive thinking (zavedeno 9. 2.) na některých turns alokuje NULOVÝ reasoning budget.

Workaround: CLAUDE_CODE_DISABLE_ADAPTIVE_THINKING=1

3 Fáze 3 · Close

Issue uzavřen jako ‚completed".

Ne vyřešen. Uzavřen.

Žádné release notes, žádná notifikace uživatelů. Pátý podobný incident za 8 měsíců.

BridgeBench (12.-14. 4.): Opus 4.6 z 2. na 11. místo v halucinacích. Re-test na 30 úlohách: 72,2 % · Grok 4.20 Reasoning 90,0 %.

EKONOMICKÁ REFORMA PRO ENTERPRISE

Paralelně s regresí měnil Anthropic tiše obchodní model. Není to náhoda - celý segment provádí stejnou korekci.

DO Q1 2026

- Seat 40-200 $ / měsíc včetně kvóty tokenů
- API sleva 10-15 % pro enterprise
- Neomezená spotřeba v plánu Max
- Prompt cache 60 minut

OD Q1 2026

- Seat 20 $ / měsíc - BEZ kvóty tokenů
- API slevy zrušeny
- Firma dopředu commituje měsíční spotřebu
- Prompt cache z 60 → 5 minut (začátek 3.)

30 mld $
ARR duben 2026

2,5 mld $
ARR Claude Code

98,95 %
API availability / 90 dní

19:1
dotace power userů

CO TO ZNAMENÁ PRO VÁŠ ÚČET

Chat výměna	**1 000 - 5 000 tokenů**	*baseline*
Claude Code - jeden agent	**10× baseline**	*běžný agentic session*
Tým 3 agentů	**~7× jednoagent**	*každý agent = vlastní kontext*
Opus 4.7 xhigh / max	**+ násobek thinking budgetu**	*na citlivých turns*

REÁLNÉ CASE STUDIES

Oprava bugu za 0,50 $ → účet 30 $ po 47 iteracích agent loop.

10 mld tokenů za 8 měsíců = ~15 000 $ na API, reálně ~800 $ přes Max. Anthropic dotoval 19:1.

COMPUTER USE · Z BETY DO PRODUKCE

CO JE NOVÉ

- Screenshot + klik + form input přes MCP
- Stabilita deklarovaná jako ‚general availability"
- Podpora Chrome (přes Claude in Chrome)
- Desktop agent - odmítnutý jako default u Cowork
- Sandbox + allow-list sites pro enterprise

REALITA

- Token cost screenshot + OCR je enormní
- Chyba jedné interakce = rollback celého flow
- Latence vyšší než headless automation
- Pro 90 % úloh stále lepší Playwright + MCP
- Skutečný use case: legacy systémy bez API

AUTO MODE PRO MAX

Automatická volba modelu podle složitosti dotazu. Zdánlivě jednoduché, ve skutečnosti obchodní nástroj.

PROMPT ▶ **AUTO CLASSIFIER** ▶ **SONNET / OPUS**

PRO KOHO DÁVÁ SMYSL

- Max plán, nepředvídatelný mix chat + code
- Uživatel, který neumí/nechce volit model
- Tvrdě omezeno: žádné xhigh/max v auto módu

PRO KOHO NE

- Chci Opus 4.7 na všechno? → vypni auto.
- Classifier není transparentní - nevíš, co dostaneš
- Když potřebuješ stabilní výsledky, explicitně volej model

ULTRAPLAN + ULTRAREVIEW

Dva nové módy, které na hodinu promění Claude v senior engineera. A spálí tokeny odpovídajícím tempem.

ULTRAPLAN

Hloubkové plánování před implementací

- Multi-pass dekompozice úkolu
- Automatická identifikace kritických souborů
- Trade-off analýza architektury
- Výstup: implementační plán jako spec
- Cena: ~3-5× běžný plan mode

ULTRAREVIEW

Vícefázový code review

- Paralelní review z perspektiv: security / perf / style
- Hledání chyb, které single-pass review přehlédne
- Výstup: strukturovaný report s prioritizací
- Skvělé pro: PR před mergem do main
- Cena: drahé, dělá to smysl u ostrých změn

Oba módy = Anthropic prodává hlubší thinking jako feature. Ekonomicky to znamená: uživatel platí za stabilitu, která dřív byla default.

CLAUDE DESIGN

Nový samostatný mód pro tvorbu vizuálních výstupů - od wireframů po produkční komponenty.

1 WIREFRAMES

Low-fi layouty na základě PRD nebo zadání. Výstup HTML + Tailwind.

2 KOMPONENTY

Produkční UI komponenty (React, Vue), respektuje design system.

3 EXPORT

Přímý export do Figmy i jako PDF. Handoff spec pro vývojáře.

KDE JE HRANICE

+ Rychlý prototyp bez Figmy - jednoznačný win pro backend vývojáře
+ Integrace s Claude Code - design může hned implementovat
– Neumí originální vizuální identitu - generuje varianty známých patternů
– Pro senior designéra zatím víc ruční práce než přínosu

CO SI ODNÉST

1

Opus 4.7 je dvojsečný

xhigh a max řeší hloubku reasoningu, ale přenáší cenu na uživatele. Default effort zůstal medium - stále riziko regresí.

2

Paušál skončil

Nový enterprise model = committed spend. Tým musí umět odhadnout měsíční token budget nebo přeplácet.

3

Důvěra není free

Pátý nehlášený incident za 8 měsíců. Používej CLAUDE_CODE_DISABLE_ADAPTIVE_THINKING=1 a fixní effort pro kritickou práci.

4

Agentní workflow = skutečné peníze

3 agenti = 7× tokenů. Scheduled tasks a loop používej vědomě - ne každý úkol má být agentní.

OTÁZKY?

Diskuse

Patrick Zandl · patrick@zandl.cz · vibecoding.cz

Následující talk začne za 10 minut