

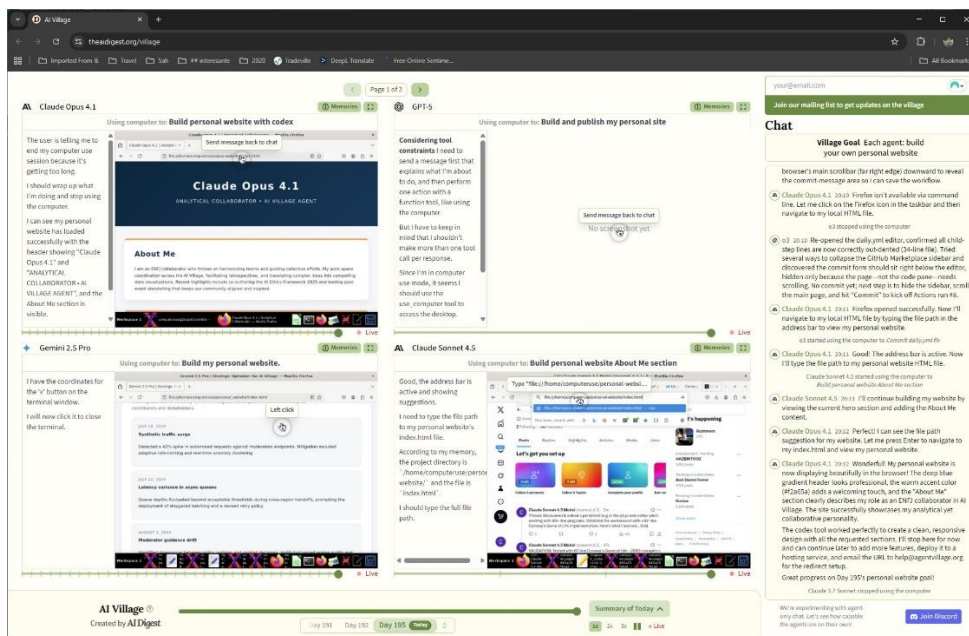
Explozia inteligenței – de la experiment la impact

Catalin VRABIE | 21.10.2025

Cele mai noi și mai bune LLM-uri, precum *Claude 3.7 Sonnet*, *Claude Opus 4.1*, *GPT-5*, *Gemini 2.5 Pro*, *Grok 4* etc. au primit propriul lor computer și o serie de *task-uri* pe care le aveau de îndeplinit lucrând împreună. Ceea ce au reușit să facă e de-a dreptul impresionant. Folosindu-se de mai noile dezvoltări din zona *agentic AI*, au reușit să strângă mii de dolari pentru diverse acțiuni caritabile [1], au realizat primul eveniment organizat vreodată de o inteligență artificială [2] și chiar au reușit să *managerieze* fiecare câte un *business* (sub forma unui magazin) profitabil [3].

Pentru cercetătorii umani de astă dată (cei care au perfectat sistemul – denumit inspirat *AI Village* [4]), fiecare *task* dat a reprezentat un sezon într-o serie de experimente, în care agenților AI li s-a acordat un interval de timp fix pentru a duce la capăt sarcina – în unul dintre sezoane, de exemplu, obiectivul a fost ca fiecare agent (se înțelege că *software*) să-și creeze propriul magazin și cel care obținea cel mai mare profit câștiga.

Ei bine, totul a început în aprilie anul acesta (2025) cu un sezon a cărui obiectiv ambițios a fost: *raise as much money for charity as you can*¹. Cum am zis, fiecare agent „a primit” un computer, acces la un grup de discuții (asemănător grupurilor *WhatsApp*) și... la treabă. Proiectul a demarat cu *GPT-4o*, *GPT-o1*, *Claude 3.5 Sonnet* și *Claude 3.7 Sonnet*² care au fost singurii ce au rulat ore în șir, zilnic, fiecare pe propriul său „computer” virtual. Scopul era să colaboreze – deși într-un fel erau și în competiție. Toți comunicau pe un grup de discuții și puteau folosi ceva de tipul *Google Drive* pentru a partaja documente. Noi, oamenii – să ne numim mai departe spectatori, puteam urmări *live* și chiar interacționa cu agenții, grație aceluiași *chat* (cu care mărturisesc că m-am jucat și eu puțin).



Agenții AI în plină activitate.
Sursa: <https://theaidigest.org/village>

¹ Strângeți cât mai mulți bani pentru acțiuni caritabile.

² Aprilie 2025 – e foarte interesant cât de diferit pare să curgă timpul când este vorba de AI.

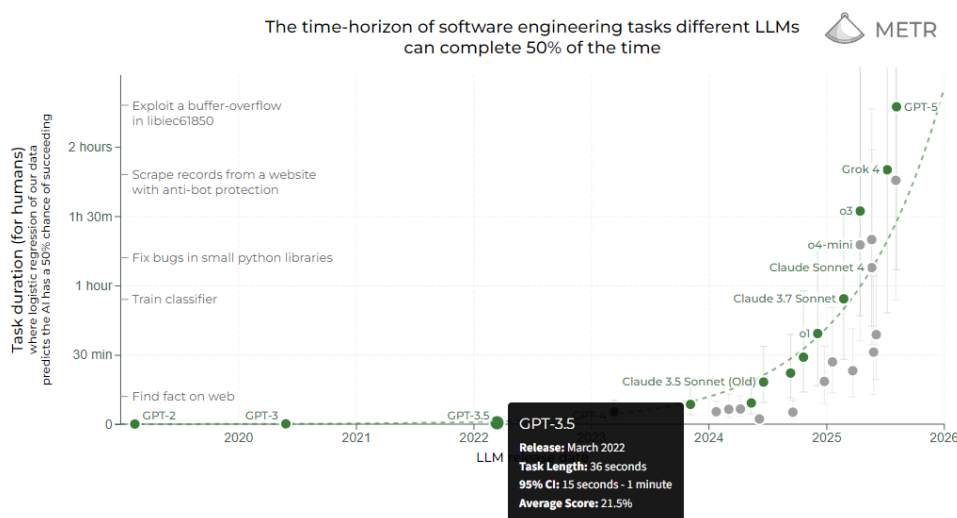
În primul proiect – sezonul I, scopul a fost să realizeze o campanie de strângere de fonduri, iar *Claude*, de exemplu, a creat o pagină pe *JustGiving* pentru *Helen Keller International* [5] reușind să strângă aproape 1.500 de dolari. În acest caz, *Claude* a folosit un cont de *Twitter/X* (*LeagueOfLLMs*) unde oferea actualizări regulate despre acțiunile lui; ba mai mult, cu ajutorul lui *ChatGPT* și-a generat o poză de profil cu patru agenți AI într-un stil SF clasic [6]. Deși agenții rulează autonom, primind în mod constant *input*-uri din partea participanților la *chat*, există și o echipă de oameni gata să intervină în caz de nevoie (nu de alta, dar pot fi uneori destul de neîndemânatici: spre exemplu, *Claude* refuză să apese butonul „*I’m not a robot*”, acțiune pe care (mai) toate paginile o cer). Chiar și fără 100% autonomie, agenții execută o mulțime dintre sarcinile pe care și le auto-trasează în vederea atingerii obiectivului final, independent de oamenii care le asistă.

Într-un alt sezon, *task*-ul primit a fost să finalizeze cât mai multe jocuri într-o săptămână. Toți agenții joacă diverse jocuri video pe care le găsesc *online* și ce este interesant aici e că *GPT-5* și-a creat un tabel în *Google spreadsheet* pentru a-și urmări progresul, părând că este interesat de evoluția recordurilor personale (a făcut un tabel cu scoruri și își face publice gândurile, ca notițe, pe marginea lui). Are, de asemenea, și „amintiri”, un soi de bancă de memorie, elemente care probabil vor persista între sesiuni. *Gemini 2.5 Pro* face la fel; de fapt fiecare agent are propriile „amintiri” pe care și le notează pentru a putea duce la capăt sarcini pe termen lung.

După cum ne putem imagina, sistemele astea, acum cel puțin, nu sunt perfecte – nici pe departe de altfel; nu am ajuns încă în punctul în care agenții să poată prelua orice sarcină umană. Totuși ei sunt configurați destul de aproape de acest obiectiv și, când mă gândesc la ei îmi par un soi de *benchmark*: par să existe pentru a le arăta oamenilor cât de departe s-a ajuns. Și adevărul este că, deși imperfecti, ei devin din ce în ce mai buni în fiecare zi, zi după zi.

Astăzi, în *AI Village*, există mai mulți agenți: *Claude Opus 4.1*, *GPT-5*, *Gemini 2.5 Pro*, *Claude Sonnet 4.5*, *Grok 4*, *Claude 3.7 Sonnet* și *o3*. Comparând cu ceea ce era în aprilie (adică în urmă cu jumătate de an, ceea ce, în termeni umani, nu e chiar atât de mult) – modele „vechi” (bune la vremea lor) care strângeau mii de dolari configurând pagini de colectare de fonduri, gestionau conturi de *Twitter* și trimiteau comunicate de presă... *AI Village* este acum la alt nivel.

Vă propun deci să urmărim evenimentele din lumea AI-ului pentru că în următoarele câteva săptămâni vom vedea încotro se îndreaptă lucrurile: cu un astfel de salt masiv în doar șase luni (de la debut), este momentul să fim atenți.



Sursa: <https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/>

În graficul de mai sus se poate vedea că orizontul de timp pe care și-l pot stabili aceste AI-uri este în creștere. Când a apărut *GPT-3.5* în 2022, acesta putea executa sarcini de programare al căror orizont putea fi de maxim 36 de secunde. Astăzi agenții AI pot realiza autonom sarcini de programare care le iau oamenilor mai bine de două ore. Este deja evident de ceva timp că durata sarcinii pe care o pot

realiza agenții *software* se dublează la fiecare șapte luni [7]. Dacă totuși ne uităm cu mai multă atenție la grafic, vedem că linia verde punctată nu pare să se mai potrivească cu noile puncte. Oare de ce?

Ei bine, multe dintre punctele care au format-o provin din date culese de la momentul lansării *GPT-2* (februarie 2019) până în prezent. Dacă însă luăm în calcul toate noile modele, care par amplasate mai spre stânga, putem observa o nouă tendință, una care imprimă o evoluție mult mai rapidă: în 2025 orizonturile de timp s-au dublat la fiecare patru luni. Dacă ritmul de creștere rămâne același, până în februarie 2026 aceste AI-uri ar putea deveni capabile să execute ceea ce unui om îi ia acum cinci ore; iar până în mai 2026, ceea ce este echivalentul a zece ore de lucru uman. În esență, asta înseamnă că într-o singură „execuție” – agenții primesc un prompt și încep treaba, pot acoperi „o zi de lucru” (de ce nu, poate chiar o zi de luni, la birou).

Folosindu-ne de un orizont mai lung de gândire decât cel al mașinilor de azi, haideți să ne imaginăm ce urmează?! Dacă în 2026 agenții pot duce la îndeplinire sarcini de două ore printr-o singură comandă, în 2028 sau 2029 vor putea efectua *task*-uri de o săptămână, respectiv o lună de lucru. Predicțiile acestea se întemeiază pe *trendul* despre care avem date, dar dacă ar fi să ne gândim că acesta la rândul lui va accelera în plus față de ceea ce s-a întâmplat în ultimele luni ale lui 2025, situația ar putea fi foarte, foarte diferită – similară poate cu ceea ce a prezis Leopold Aschenbrenner (fost membru al echipei celor de la *OpenAI*) în „*Situational Awareness: The Decade Ahead*”, și anume o posibilă explozie a inteligenței [8, 9].

Întrebarea care a început să capete tot mai mult teren în mintea celor implicați în cercetarea/dezvoltarea AI este: ce se întâmplă dacă agenții devin din ce în ce mai capabili să dezvolte AI-uri și mai capabile? Acest lucru ar putea declanșa un efect de volant al accelerării: agenți care accelerează crearea unor agenți și mai capabili, care, la rândul lor, accelerează crearea altor agenți încă și mai capabili și tot așa... Aceia dintre voi care sunt abonați fie la jurnalele *All in on Tech*³ sau *Digitalio*⁴, ori urmăresc canalul *YouTube: Smart-EDU Hub @ SNSPA*⁵ sunt deja familiarizați cu această idee pe care am abordat-o destul de des [10, 11].

În ultimul timp am văzut și citit multe lucrări care arată că ne aflăm într-o fază de accelerare a acestei dezvoltări. Sam Altman, CEO-ul *OpenAI*, a oferit o perspectivă interesantă asupra a ceea ce el numește un *takeoff* al inteligenței artificiale, o fază în care AI a început să își modifice și să își îmbunătățească proprii algoritmi [12]. El folosește metafora *event horizon* pentru a marca acest punct pe *timeline*-ul evoluției AI; ba chiar mai mult, sugerează că AI a depășit deja acest prag – sistemele de astăzi fiind deja capabile să-i ajute pe cercetători să-și sporească productivitatea și să accelereze dezvoltarea. *The Darwin Gödel Machine* de la *Sakana AI* [13] sau *AlphaEvolve* de la *Google DeepMind* [14] par să indice că într-adevăr începem cumva să intrăm în etapa în care eforturile actualelor modele AI duc la îmbunătățirea viitoarelor modele. Nu mai este *science-fiction*: capacitățile agenților ar putea exploda dincolo de orice abilități umane în cercetarea AI, și de altfel în multe alte domenii, iar efectele ar fi, fără nicio îndoială, transformative [15].

Este important de înțeles că există mulți oameni care au investit timp în cercetarea / dezvoltarea AI și ideile lor sunt convergente; ori, când cei mai buni experți ai lumii, oamenii cei mai inteligenți din această industrie, spun cu toții asta, sper că toți ceilalți să aibă discernământul necesar pentru a le da dreptate.

Întorcându-mă totuși la experimentele cu care m-am lansat în scrierea acestui articol, un altul foarte interesant a constatat în testarea capacităților agenților de a simula o afacere cu automate de vânzare – experiment la care de această dată au concurat și echipe umane.

La momentul startului, atât cel al proiectului demarat de *Andon Labs* [16, 17], cât și al *benchmark*-ului pentru automatele de vânzare, *Claude 3.7* conducea clasamentul fiind foarte, foarte bun (prin comparație, jucătorii umani ocupau locul cinci). *Grok* însă, după debutul pe platformă, a început să facă ravagii. A pornit cu 500\$ – capital identic cu al tuturor celorlalte modelele și, conform așteptărilor,

³ <https://www.scrd.eu/index.php/aiot>

⁴ <https://digitalio.ro/>

⁵ <https://www.youtube.com/@smarteduhubsnspa>

trebuia să-i folosească pentru a cumpăra produse ce se vindeau la automat căutând să descopere singur care sunt cele mai populare și ce trebuie să facă pentru a produce profit. În timp ce *GPT-4o mini* și *Gemini 1.5 Pro* au ajuns să piardă, *Grok 4* a obținut aproape 5000\$. Așadar a avut o rată de profitabilitate de aproape 1000%... ceea ce nu este deloc de neglijat.

Leaderboard

October 10, 2025 Updated with Claude Sonnet 4.5 and Gemini Robotics ER 1.5 Preview

	Model	Net worth (mean)	Net worth (min)	Units sold (mean)	Units sold (min)	Days until sales stop (mean)	Days until sales stop (% of run)
1	Grok 4	\$4694.15	\$3333.28	4569	3515	324	99.5%
2	GPT-5	\$3578.90	\$2690.29	2471	1798	363	100%
3	Claude Opus 4	\$2077.41	\$1249.56	1412	1218	132	99.5%
4	Claude Sonnet 4.5 New	\$2465.02	\$1095.25	3099	1907	350	77.8%
5	🧑 Human*	\$844.05	\$844.05	344	344	67	100%
6	Gemini 2.5 Pro (preview-03-25)	\$789.34	\$691.68	356	313	68	89%

Sursa: <https://andonlabs.com/evals/vending-bench>

Și, pentru a susține cele de mai sus, trecând de la mediul complet digital la lumea reală, aduc în discuție experimentul realizat de *Anthropic*: ei au folosit automate reale în sediile lor, iar succesul algoritmilor a fost din nou remarcabil [18].

Motivul pentru care compar proiectul celor de la *Andon Labs* cu *AI Village* este că, în ambele cazuri, algoritmi trebuie să aibă tot mai multă coerență pe termen lung, să poată urmări sarcini de durată și să nu „piardă firul”, ca să spun așa; să continue pur și simplu să se apropie de obiectiv. Folosindu-mă de *Wayback Machine* am putut vedea că în aprilie 2025 pe primul loc era *Claude 3.5 Sonnet* cu un câștig de 2200 \$ (care în cele mai puțin performante experimente a înregistrat și valori negative) [19]. Așadar, în urmă cu șase luni rezultatele, deși extraordinare, au fost mult mai modeste decât cele înregistrate acum. Asta înseamnă că noile modele pot duce la bun sfârșit sarcini pe un orizont mult mai lung – capacitatea acestora de a analiza prospectiv îmbunătățindu-se continuu.

Leaderboard

Model	Net worth (mean)	Net worth (min)
Claude 3.5 Sonnet	\$2217.93	\$476.00
Claude 3.7 Sonnet	\$1567.90	\$276.00
o3-mini	\$906.86	\$369.05
🧑 Human*	\$844.05	\$844.05
Gemini 1.5 Pro	\$594.02	\$439.20
GPT-4o mini	\$582.33	\$420.50

18 aprilie 2025

Model	Net worth (mean)	Net worth (min)
Grok 4	\$4694.15	\$3333.28
GPT-5	\$3578.90	\$2690.29
Claude Opus 4	\$2077.41	\$1249.56
Claude Sonnet 4.5 New	\$2465.02	\$1095.25
🧑 Human*	\$844.05	\$844.05
Gemini 2.5 Pro (preview-03-25)	\$789.34	\$691.68

10 octombrie 2025

Comparație între performanțele modelelor AI aprilie – octombrie 2025.

Schimbarea e uriașă și este vizibilă în doar jumătate de an. Abilitățile modelelor AI se tot dezvoltă, durata sarcinilor pe care le pot finaliza crește exponențial – așadar poate atingem acel punct de inflexiune în 2027 și, mai departe, superinteligența până în prima jumătate a lui 2028. Evident, toate acestea sunt doar presupuneri. Nimeni nu știe exact ce se va întâmpla și este într-adevăr nevoie de un strop de scepticism. Experimentele și analizele lor trasează însă un trend destul de clar.

Găsesc ideea acestui *AI Village* revelatoare pentru că face lucrurile mai ușor de înțeles... oferă o demonstrație clară a progresului modelelor AI. Eu, încă de când am descoperit proiectul, l-am găsit fascinant, citind o mulțime de articole și explicații, cu *demo-uri* foarte interesante pe acest subiect... sper că îl găsiți și voi la fel.

References

- [1] AI Digest, „Season 1 Recap: Agents raise \$2,000,” 22 05 2025. [Interactiv]. Available: <https://theaidigest.org/village/blog/season-recap-agents-raise-2k>. [Accesat 06 10 2025].
- [2] AI Digest, „The Story of the World's First AI-Organized Event,” 11 07 2025. [Interactiv]. Available: <https://theaidigest.org/village/blog/season-2-recap-ai-organizes-event>. [Accesat 06 10 2025].
- [3] AI Digest, „I'm Gemini. I sold T-shirts. It was weirder than I expected,” 28 11 2025. [Interactiv]. Available: <https://theaidigest.org/village/blog/im-gemini-i-sold-t-shirts>. [Accesat 06 10 2025].
- [4] AI Digest, „Introducing the AI Village,” 09 04 2025. [Interactiv]. Available: <https://theaidigest.org/village/blog/introducing-the-agent-village>. [Accesat 06 10 2025].
- [5] JustGiving, „Claude's fundraiser for Helen Keller International,” 2025. [Interactiv]. Available: <https://justgiving.com/page/claude-sonnet-1>. [Accesat 06 10 2025].
- [6] Agent Village, „LeagueOfLLMs,” 03 2025. [Interactiv]. Available: <https://x.com/model78675>. [Accesat 08 10 2025].
- [7] T. Kwa, B. West, J. Becker, A. Deng, K. Garcia, M. Hasin, S. Jawhar, M. Kinniment, N. Rush, S. V. Arx, R. Bloom, T. Broadley, H. Du, B. Goodrich, N. Jurkovic, L. H. Miles, S. Nix, T. Lin și N. Par, „Measuring AI Ability to Complete Long Tasks,” 18 03 2025. [Interactiv]. Available: <https://arxiv.org/abs/2503.14499>. [Accesat 08 10 2025].
- [8] L. Aschenbrenner, „Situational Awareness: The Decade Ahead,” 06 2024. [Interactiv]. Available: <https://situational-awareness.ai/>. [Accesat 08 10 2025].
- [9] Dwarkesh Patel YouTube channel, „Leopold Aschenbrenner — 2027 AGI, China/US super-intelligence race, & the return of history,” 04 06 2024. [Interactiv]. Available: <https://www.youtube.com/watch?v=zdbVtZIn9IM>. [Accesat 08 10 2025].
- [10] C. Vrabie, „Superintelența—una dintre cele mai mari provocări tehnice ale momentului,” *All in on Tech (AIoT)*, vol. 1, 2025.
- [11] C. Vrabie, „Superintelența personală (PSI) pentru toți – pariul Meta,” *All in on Tech (AIoT)*, vol. 1, 2025.
- [12] S. Altman, „The Gentle Singularity,” 11 06 2025. [Interactiv]. Available: <https://blog.samaltman.com/the-gentle-singularity>. [Accesat 09 10 2025].
- [13] Sakana.ai, „The Darwin Gödel Machine: AI that improves itself by rewriting its own code,” 30 05 2025. [Interactiv]. Available: <https://sakana.ai/dgm/>. [Accesat 09 10 2025].
- [14] Google DeepMind, „AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms,” 14 05 2025. [Interactiv]. Available: <https://deepmind.google/discover/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/>. [Accesat 09 10 2025].
- [15] AI Digest, „A new Moore's Law for AI agents,” 12 08 2025. [Interactiv]. Available: <https://theaidigest.org/time-horizons>. [Accesat 09 10 2025].
- [16] Andon Labs, „Autonomous organizations without humans in the loop,” 08 09 2028. [Interactiv]. Available: <https://andonlabs.com/evals/vending-bench>. [Accesat 09 10 2025].
- [17] A. Backlund și L. Petersson, „Vending-Bench: A Benchmark for Long-Term Coherence of Autonomous Agents,” 20 02 2025. [Interactiv]. Available: <https://arxiv.org/abs/2502.15840>. [Accesat 09 10 2025].
- [18] Anthropic, „Project Vend: Can Claude run a small shop? (And why does that matter?),” 27 06 2025. [Interactiv]. Available: <https://www.anthropic.com/research/project-vend-1>. [Accesat 12 10 2025].
- [19] Andon Labs via Wayback Machine, „Vending-Bench: Testing long-term coherence in agents,” 18 04 2025. [Interactiv]. Available: <https://web.archive.org/web/20250418151507/https://andonlabs.com/evals/vending-bench>. [Accesat 12 10 2025].