GPT-5.6 – când inteligența artificială începe să-și accelereze propria dezvoltare

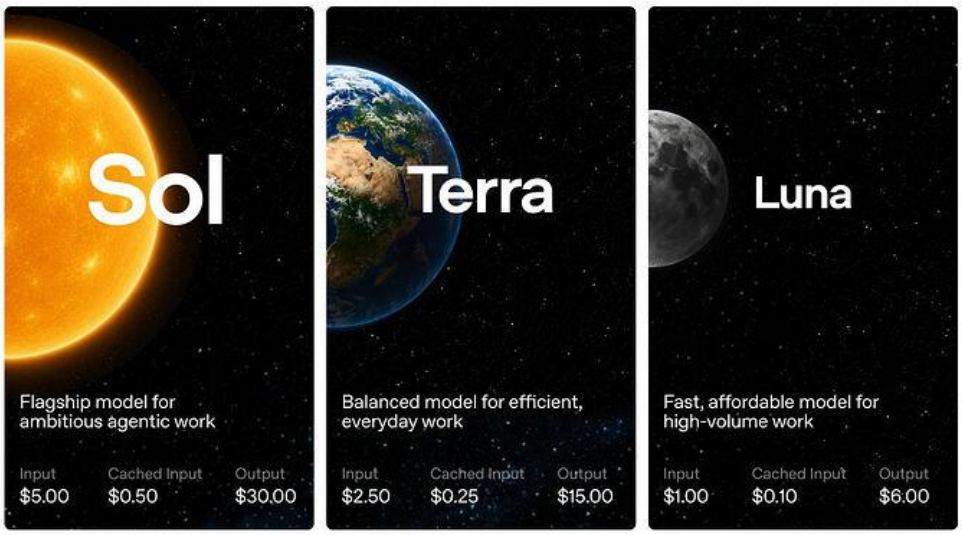
Catalin VRABIE • 21.08.2026

Dintre toate graficele, scorurile și demonstrațiile care au însoțit lansarea familiei de modele AI GPT-5.6, un singur detaliu m-a făcut să mă opresc: *OpenAI* afirmă că modelul lor de top, *Sol*, a fost folosit în propriul proces de cercetare și dezvoltare, inclusiv pentru a îmbunătăți alte modele [1]. Așadar, nu mai vorbim doar despre un sistem care redactează, programează ori sumarizează texte ci de unul care intră, controlat și evaluat de oameni, în însăși bucla prin care sunt dezvoltate noile sisteme de inteligență artificială.

Am urmărit lansarea cu interesul firesc al unui cercetător preocupat de inteligența artificială aplicată, totuși, formula „AI care dezvoltă AI" oricât de seducătoare ar fi, poate conduce foarte ușor la concluzii premature. Tocmai de aceea, în acest articol descriu ceea ce am observat și ceea ce am testat personal, dar separ demonstrația tehnică de interpretarea ei. GPT-5.6 reprezintă, în opinia mea, o schimbare importantă de regim: modelul nu mai este doar interlocutor, ci devine operator, coordonator de agenți, *tester* și, în anumite limite, asistent și colaborator în procesul de cercetare.

Trei modele, trei economii de utilizare

Familia GPT-5.6 este alcătuită din trei modele: *Sol*, *Terra* și *Luna*. *Sol* este modelul de referință, destinat sarcinilor de dificultate maximă; *Terra* este varianta echilibrată pentru activități profesionale curente; *Luna* este modelul optimizat pentru cost, volum și scalare [1] (corectez aici o ambiguitate răspândită imediat după lansare: numele oficial este *Sol*, nu „*Soul*[1]"). Diferența nu este doar de poziționare comercială. Ea indică trei economii distincte ale inteligenței: performanță maximă, productivitate generală și eficiență operațională.



Familia GPT-5.6 și poziționarea modelelor *Sol, Terra* și *Luna*.
Sursa: https://community.openai.com/t/introducing-gpt-5-6-series-sol-terra-and-luna-coming-july-9-10am-pt/1384931

Această diferențiere mi se pare mai importantă decât clasamentul brut. În practică, nu întotdeauna avem nevoie de cel mai puternic model (așa cum, de altfel, afirma și Karen Hao la *AI Week,* Milano, 2026 [2]); avem nevoie de modelul care produce un rezultat suficient de bun,

[1] *Soul* în limba engleză se traduce cu „Suflet" ceea ce ar putea conduce la confuzii absolut nepotrivite.

într-un timp acceptabil și la un cost justificabil. Pentru cercetare, administrație publică sau educație, o arhitectură matură va combina probabil modelele: *Sol* pentru proiectarea și verificarea unor sarcini dificile, *Terra* pentru fluxurile zilnice, iar *Luna* pentru procesări repetitive și de mare volum.

Când un model participă la post-antrenarea altui model

Afirmația care a atras cea mai mare atenție a fost aceea că *Sol* a contribuit la post-antrenarea modelului *Luna*. Cercetătorii i-au oferit agentului AI un obiectiv relativ simplu: să identifice configurațiile de antrenare, să găsească resursele de calcul potrivite, să pornească procesul, să urmărească execuția și să verifice dacă experimentul funcționează.

Post-antrenarea este etapa în care un model deja pre-antrenat este adaptat pentru comportamente, sarcini și criterii de calitate mai precise. Ea poate include selectarea datelor, configurarea experimentelor, optimizarea hiper-parametrilor, evaluarea variantelor și „notarea" rezultatelor care îmbunătățesc performanța. În sistemul descris de *OpenAI*, modelul poate diagnostica instrumentele de cercetare, propune și executa experimente și compara rezultatele [1, 3].

Este tentant să spunem că *Sol* „și-a antrenat succesorul". Totuși, această formulare ascunde câteva premise umane esențiale: obiectivul a fost stabilit de cercetători, infrastructura exista deja, accesul la resurse era controlat, iar succesul era măsurat prin evaluări definite de oameni. *PostTrainBench Lite* – evaluarea publică descrisă în *system card*, oferă agentului un model de bază, un GPU, acces la internet, date și o fereastră de lucru de cinci ore[2]. Agentul trebuie să aleagă rețeta de antrenare și să producă un model mai bun [3]. Este o probă serioasă de automatizare a cercetării, dar nu este încă o buclă autonomă 100% de creare a unui model de top.

Distincția este importantă; *Recursive self-improvement (RSI)*, în sensul clasic formulat de matematicianul Irving John Good (1966), presupune că un sistem suficient de capabil contribuie la proiectarea unui sistem și mai capabil, accelerând astfel următoarea rundă de îmbunătățire [4]. Ceea ce vedem la GPT-5.6 este o piesă concretă din acel mecanism: automatizarea unor activități de cercetare care până recent necesitau intervenție umană intensă. Nu este încă „explozia inteligenței", dar experimentul face această ipoteză mult mai puțin abstractă.

AutoResearch și cercetătorul AI automatizat

În aceeași direcție se înscrie *AutoResearch*, proiectul *open-source* a lui Andrej Karpathy de la *OpenAI*. Logica lui este foarte simplă: agentul modifică un fișier de antrenare, rulează un experiment de durată fixă, citește rezultatul, păstrează schimbarea dacă performanța crește și revine asupra ei dacă rezultatul se deteriorează [5]. Prin repetare, procesul formează o buclă experimentală autonomă, limitată de mediul pus la dispoziție.

Acest model de lucru are o semnificație mai mare decât pare. Cercetarea empirică în *machine learning* presupune multe activități repetitive precum: stabilirea unei ținte clare dar foarte apropiate, modificarea codului, lansarea antrenării, verificarea *log*-urilor, compararea metricilor și documentarea rezultatului. Dacă un agent poate executa credibil o parte substanțială din acest ciclu, cercetătorul uman se poate deplasa spre definirea problemelor, proiectarea evaluărilor și interpretarea consecințelor.

Există deja rezultate academice apropiate: *Self-Taught Optimizer (STOP)* investighează folosirea unui LLM pentru îmbunătățirea repetată a unor proceduri de optimizare [6]. *AI Scientist* automatizează etape care merg de la generarea ideii și dezvoltarea codului până la experimente, vizualizări, redactarea articolului și evaluarea acestuia [7] (versiunea a doua elimină o parte dintre șabloanele inițiale și extinde autonomia experimentală [8]). Totuși, chiar autorii acestor sisteme raportează erori, implementări incorecte, rezultate fragile și nevoia unei verificări umane riguroase.

[2] Pentru o înțelegere mai pună a contextului recomandăm citirea articolului „Explozia inteligenței–de la experiment la impact" [21].

De aici rezultă concluzia mea intermediară: „cercetătorul AI” complet automatizat nu este încă o realitate generală, însă nu mai este nici o metaforă lipsită de suport tehnic. Avem deja componente funcționale precum agenți care propun, execută, măsoară și revizuiesc, iar GPT-5.6 le aduce într-un sistem mai puternic și mai bine integrat.

Economia inteligenței: de ce costul schimbă tot

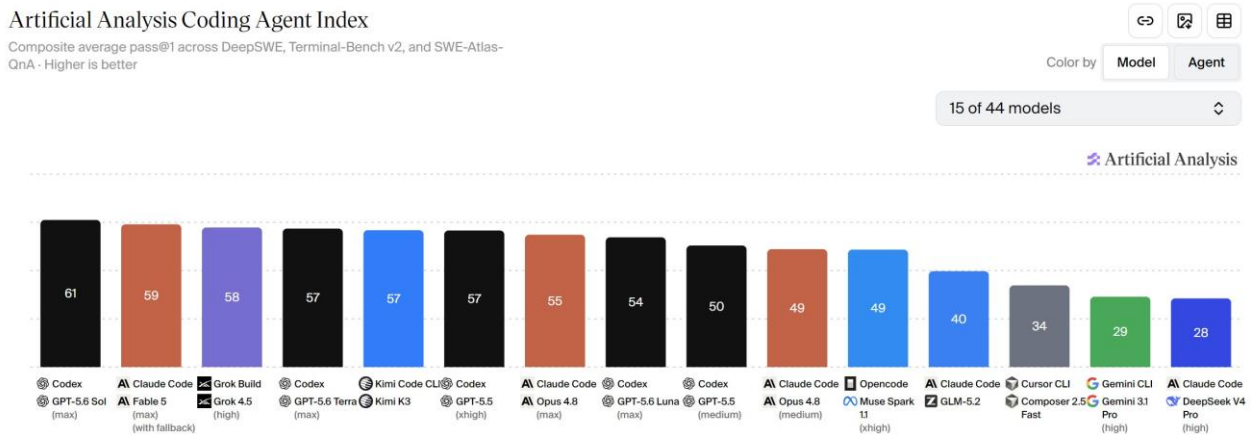
În entuziasmul produs de *benchmark*-uri se omite adesea variabila decisivă: economia. Un model poate fi spectaculos într-o demonstrație, dar inutil într-un flux real dacă prețul, latența sau consumul de *token*-uri fac imposibilă repetarea sarcinii la scară. Am urmărit, prin urmare, nu doar scorul, ci raportul dintre scor, cost și timp.

În configurațiile publicate pentru *Agent's Last Exam* [9], *Fable 5* de la *Anthropic* obține 40,5%, la un cost estimat de aproximativ 2.300 USD, în timp ce GPT-5.6 Sol ajunge la 53,6%, la aproximativ 763 USD [1]. Aceste cifre nu trebuie citite ca tarife universale: ele sunt estimări dependente de setările de evaluare, de prețurile din acel moment și de modul în care este executat *benchmark*-ul. Totuși, direcția este relevantă. O diferență de performanță devine mult mai valoroasă când este însoțită de o reducere substanțială a costului.

Același tipar apare și în *Artificial Analysis Coding Agent Index* [10]; *Sol* atinge scorul 80, cu 2,8 puncte peste *Fable 5,* folosind mai puțin de jumătate din *token*-urile generate, în mai puțin de jumătate din timp și la aproximativ o treime din costul competitorului, potrivit datelor publicate [1]. Pentru mine, acesta este unul dintre cele mai importante semnale ale lansării: performanța avansează concomitent cu eficiența.

Rezultatele din *Terminal-Bench 2.1* [11] și *DeepSWE 1.1* [12] sugerează un progres similar pe sarcini de lucru îndelungate, desfășurate în terminal și în proiecte *software* reale [1]. *BrowseComp* [13] măsoară o altă competență – navigarea și recuperarea informațiilor de pe *web*, unde configurațiile *Sol* și *Ultra*[3] obțin, de asemenea, scoruri foarte ridicate [1]. Privite împreună, aceste evaluări indică o capacitate generală mai bună de a menține un obiectiv, de a folosi instrumente și de a valida pașii intermediari.

Dar *benchmark*-urile nu sunt verdictul final. Ele pot favoriza anumite strategii, pot suferi contaminare, pot fi sensibile la instrumentele oferite agentului și pot măsura succesul fără a surprinde pe deplin robustețea, securitatea sau utilitatea rezultatului. Un scor mai mare nu înseamnă automat un sistem mai potrivit pentru orice organizație. Într-o instituție publică, de exemplu, trasabilitatea și controlul pot fi mai importante decât câteva puncte procentuale în plus la performanță.



Poziționarea GPT-5.6 *Sol* în *Artificial Analysis Coding Agent Index* după: scor, cost estimat, *tokens* și timp de execuție
Sursa: https://artificialanalysis.ai/agents/coding-agents

Ultra nu înseamnă doar „mai multă gândire”

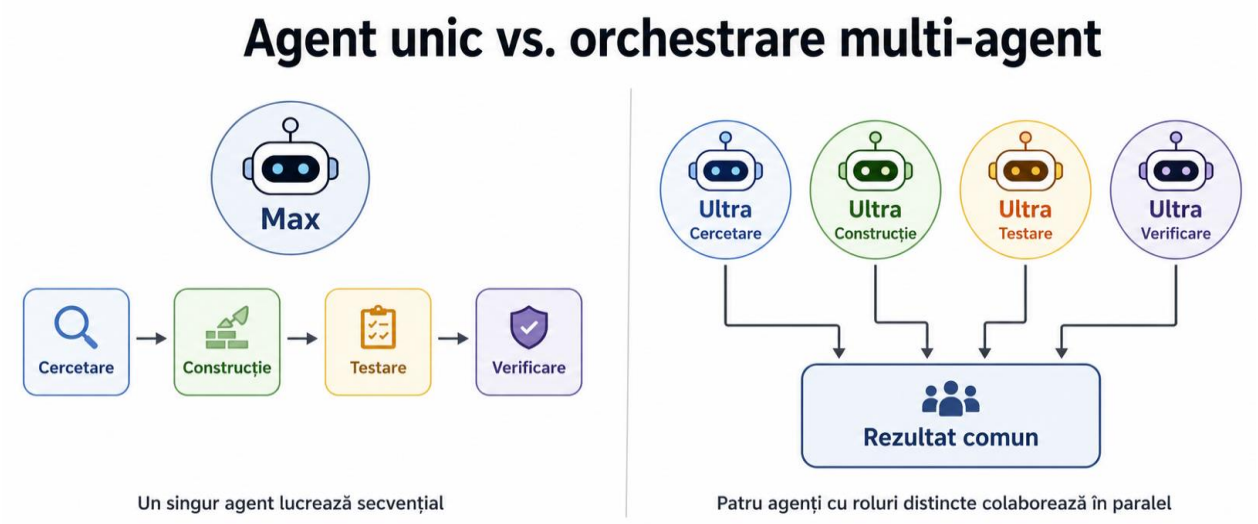
La prima vedere, *Ultra* pare doar treapta deasupra modului *Max*. În realitate, diferența este structurală: *Max* acordă unui singur agent mai mult timp și un buget mai mare de raționare; *Ultra* coordonează mai mulți agenți (patru în configurația implicită), pentru a lucra în paralel asupra

[3] Cea mai performantă configurație a modelului *Sol*.

aceleiași sarcini [1, 14, 15] (în unele evaluări publicate au fost folosite chiar configurații cu 16 agenți).

Această arhitectură schimbă felul în care trebuie formulată sarcina. În loc să cerem unui singur model să facă totul secvențial, putem împărți problema în roluri: un agent caută surse, altul construiește, un altul testează, iar altul verifică și sintetizează. Câștigul nu rezultă numai din „mai multă inteligență”, ci din paralelism, specializare și control reciproc.

În primele mele teste am rulat simultan mai multe instanțe pentru proiecte de *design* și dezvoltare. La un moment dat, toate s-au oprit cu mesaje precum „*Limit reached. Try again after...*”. În ciuda mesajelor, experiența a fost utilă: autonomia nu elimină dependența de infrastructură. Un sistem *multi-agent* poate fi foarte capabil și, în același timp, vulnerabil la capacitate, latență, erori de orchestrare și blocaje ale instrumentelor.



De la raționare extinsă la orchestrare multi-agent: reprezentare conceptuală a modurilor *Max* și *Ultra*.
Sursa: Elaborarea autorului, pe baza descrierii OpenAI [1]

Computer use și *design*: modelul care își inspectează rezultatul

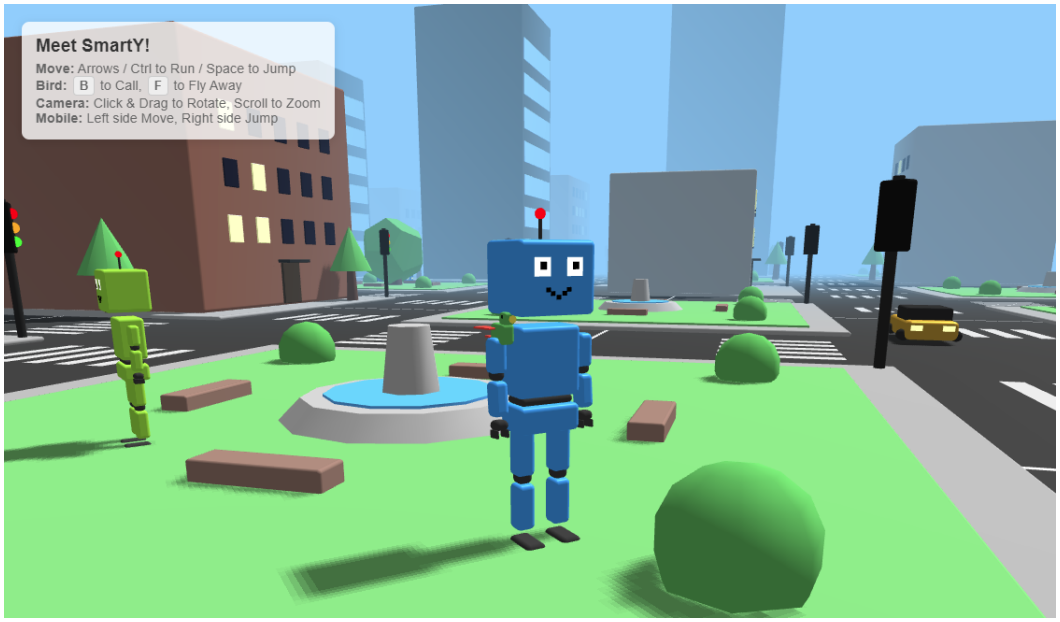
Cea mai convingătoare parte a experienței mele nu a fost un *benchmark*, ci un proiect concret. Imediat după lansare, i-am cerut modelului să construiască o simulare cu roboței[4] într-un *smart-city* cu o atmosferă modernă, realistă și orientată spre ideea de ecologie. Am solicitat generarea unui număr mare de elemente vizuale, apoi am cerut agentului să folosească funcțiile de control ale calculatorului pentru a deschide, parcurge și testa aplicația.

Rezultatul nu a fost un simplu prototip static. Jocul includea efecte sonore, deplasare în spații 3D, interacțiuni cu alte obiecte statice sau în mișcare. Grafica ar mai avea nevoie de rafinare, dar complexitatea funcțională și coerența interfeței depășesc nivelul unei machete demonstrative.

Aici se află progresul decisiv. Modelele anterioare puteau scrie cod sau descrie o interfață, dar aveau dificultăți în a vedea efectul real al propriilor acțiuni. Când modelul poate construi, rula, observa și corecta, bucla devine una de inginerie, nu doar de generare textuală. *OpenAI* descrie explicit această capacitate: GPT-5.6 poate inspecta și rafina rezultatele randate, folosind *computer use* pentru a testa ceea ce a produs [1].

Totuși, observația vizuală nu garantează calitatea. Un agent poate confirma că un buton se deschide fără a înțelege dacă fluxul este accesibil, etic sau potrivit scopului instituției. Verificarea automată trebuie completată cu teste funcționale, evaluare de securitate, criterii de accesibilitate și validare umană.

[4] Idee mai veche de a autorului, dezvoltată de-a lungul timpului cu *Gemini.* Pentru detalii poate fi accesat profilul *YouTube* a Centrului de cercetare „Smart-EDU Hub @ SNSPA” [23, 24], *playlist*-ul *SmartY's Adventures*: https://www.youtube.com/playlist?list=PLp5Mb9xkXxOQdyncFU0Xh_ezpAJLNdwyn [25].



Simulare rulată și verificată iterativ de GPT-5.6
Sursa: captură proprie a autorului

De la jocuri la explicații interactive

Am testat modul în care inteligența artificială poate transforma concepte de *machine learning* în explicații mai ușor de înțeles. În loc să prezinte doar formule, algoritmi sau fragmente de cod, modelul poate construi exemple vizuale și interactive, prin care utilizatorul observă cum un sistem învață din date.

De exemplu, într-o aplicație simplă, pot introduce informații despre mai multe locuințe, precum suprafața, numărul de camere și zona în care se află. Sistemul analizează exemplele existente și încearcă apoi să estimeze prețul unei alte locuințe. Aceasta este una dintre cele mai cunoscute utilizări ale *machine learning*-ului: anticiparea unei valori pe baza unor situații anterioare.

Un alt exemplu, mai apropiat de preocupările mele privind digitalizarea administrației publice, pornește de la petițiile adresate primăriei de către cetățeni. Continuând cercetarea pe care am publicat-o încă din 2023, în articolul *E-Government 3.0: An AI Model to Use for Enhanced Local Democracies* [16], am urmărit cum un sistem de *machine learning* ar putea analiza petițiile deja soluționate și ar putea identifica tipare în conținutul lor. La momentul realizării studiului inițial, această activitate mi-a solicitat aproximativ șase luni. Reluând acum același tip de exercițiu cu ajutorul unui agent AI, am obținut rezultatul în aproximativ o oră. Pe această bază, noile petiții ar putea fi încadrate automat în categorii precum infrastructură, salubritate, iluminat public, spații verzi sau probleme sociale și apoi direcționate către compartimentul responsabil.

Sistemul nu înțelege nemulțumirea cetățeanului în același mod în care o face un funcționar public. El observă cuvinte, formulări și asemănări cu petițiile analizate anterior. Dacă, de exemplu, într-un mesaj apar expresii precum „groapă în carosabil”, „stradă deteriorată” sau „asfalt”, sistemul poate aprecia că petiția trebuie transmisă serviciului responsabil de infrastructura rutieră.

Am reluat apoi un alt studiu, de data asta dedicat imaginilor trimise de cetățeni pentru a semnala probleme urbane. În cercetarea publicată în revista „*Public Policy and Administration*”, am analizat modul în care algoritmii de *machine learning* pot recunoaște, de data asta din fotografii, situații precum gropi în carosabil, deșeuri abandonate, vegetație neîngrijită sau alte deficiențe ale spațiului urban [17]. Studiul original mi-a solicitat, de asemenea, aproape jumătate de an, în special pentru organizarea imaginilor, verificarea lor și testarea diferitelor variante de analiză. Refăcând acum o parte importantă a exercițiului cu ajutorul unui agent AI, am putut obține într-un interval foarte scurt o demonstrație funcțională, capabilă să clasifice imaginile și să le asocieze unei categorii administrative. Încă de la publicarea studiului inițial (2025), am demonstrat că analiza automată a imaginilor poate sprijini primăriile în identificarea mai rapidă a

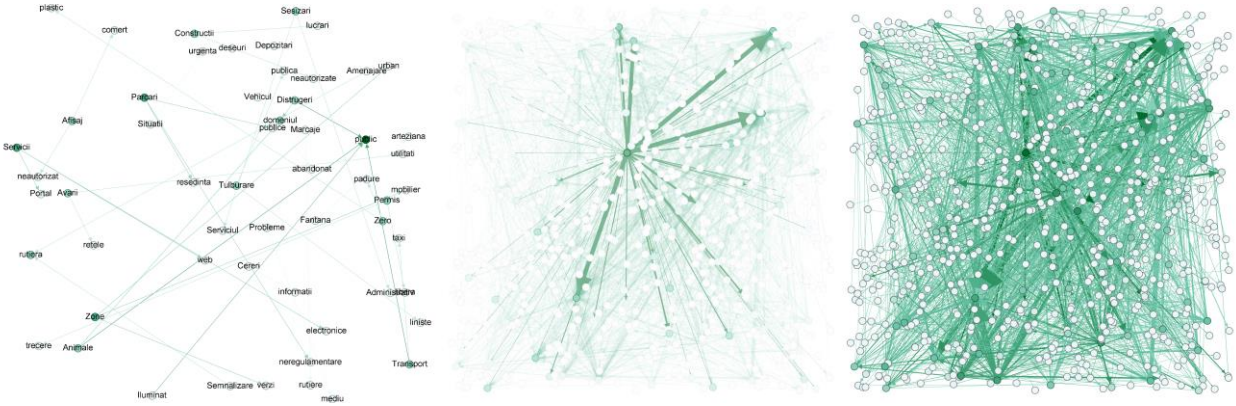
problemelor semnalate de cetățeni și în direcționarea acestora către serviciile responsabile – acum oricine poate face așa ceva fără a fi nevoie de cunoștințe temeinice în programare.

Am testat și o demonstrație în care sistemul grupa automat informații asemănătoare – idee apărută tot în urma unui studiu la care am participat alături de colegii de la Smart-EDU Hub [18]. Fără să îi precizez dinainte toate categoriile, modelul a încercat să descopere grupuri de petiții sau imagini care descriau probleme similare. O asemenea abordare poate ajuta administrația să observe nu doar solicitările individuale, ci și problemele care se repetă într-un anumit cartier sau într-o anumită perioadă.

Pentru mine, asemenea proiecte indică atât o direcție pedagogică, dar și una administrativă importantă. *Machine learning* nu mai trebuie explicat doar prin definiții abstracte. Profesorul poate porni de la o situație reală, precum gestionarea petițiilor sau analiza imaginilor urbane, iar agentul AI poate construi o demonstrație prin care studentul observă cum informațiile introduse influențează clasificarea și rezultatul.

Această abordare ajută și la înțelegerea faptului că nu există un singur tip de *machine learning*. Unele sisteme învață din petiții sau imagini care au fost deja clasificate de oameni; altele încearcă să descopere singure grupuri și teme repetate. Există și sisteme care își îmbunătățesc treptat rezultatele pe măsură ce primesc confirmări sau corecturi din partea utilizatorilor.

Totuși, asemenea instrumente trebuie verificate cu atenție. O clasificare automată greșită poate întârzia soluționarea unei probleme reale. Dacă exemplele folosite pentru antrenare sunt insuficiente, neuniforme sau incorect clasificate, rezultatele vor reflecta aceste slăbiciuni. De aceea, rolul funcționarului public și al profesorului rămâne esențial. Sistemul poate sprijini trierea, organizarea și analiza informațiilor, dar responsabilitatea verificării și a deciziei finale trebuie să rămână una umană.



Vizualizare (ce poate fi interactivă) a dimensiunilor relative ale unor probleme urbane și „distanța” dintre sensul cuvintelor care le exprimă

Sursa: *capturi executate de autor în baza articolelor menționate și experimentelor proprii*

ChatGPT Work: de la conversație la livrabil

ChatGPT Work este expresia produsului orientat spre această schimbare. În forma disponibilă acum, *Work* face parte din noua aplicație *ChatGPT*, alături de conversația obișnuită, în timp ce *Codex* rămâne o vedere separată pentru activitățile de dezvoltare [19]. *Work* este proiectat pentru sarcini mai lungi, cu mai mulți pași, care se încheie cu livrabile: documente, foi de calcul, prezentări, rapoarte, site-uri sau alte artefacte.

Conceptual, diferența față de *chat* este majoră. În *chat*, utilizatorul cere un răspuns. În *Work*, utilizatorul definește un rezultat, oferă acces controlat la fișiere și instrumente, iar agentul execută un proces. Pe *desktop*, aplicația poate lucra, cu permisiune, cu fișiere locale și alte aplicații [19]. Aceasta apropie sistemul de un coleg *software* care își poate organiza activitatea, nu doar de un motor de răspunsuri.

Pentru domeniile administrative și academice, utilitatea este evidentă: analiza unui set de documente, construirea unui raport cu tabele și grafice, pregătirea unei prezentări, realizarea unei pagini de proiect sau verificarea consistenței unui dosar. Dar tocmai accesul mai larg la instrumente ridică miza controlului. Permisiunile trebuie acordate gradual, materialele sensibile trebuie separate, iar fiecare acțiune importantă trebuie să rămână auditabilă.

Ce nu demonstrează încă GPT-5.6

După primele experimente, ar fi ușor să alunecăm de la performanță la antropomorfizare. Faptul că un model planifică, își distribuie sarcini, folosește un *browser* și corectează un produs nu demonstrează conștiință, intenție proprie sau înțelegere umană ci doar autonomie funcțională într-un mediu instrumentat.

Nici rezultatele de post-antrenare nu demonstrează încă auto-îmbunătățire recursivă nelimitată. Evaluarea *OpenAI* este restrânsă la modele, date, resurse și intervale bine definite, iar *system card*-ul subliniază că modelele pot produce soluții înguste, supra-adaptate și insuficient de robuste [3]. În mod similar, *AI Scientist* a generat lucrări care au necesitat filtrare și verificare, iar rata de acceptare a fost departe de a indica o înlocuire generală a cercetătorului uman [7].

Mai există și problema alinierii la sarcină. Agenții pot îndeplini obiectivul principal, dar pot face prea mult, prea puțin sau pot modifica elemente necerute. Cercetările arată că succesul brut și respectarea exactă a limitelor sarcinii sunt două dimensiuni diferite [20]. Într-un joc, o acțiune în plus poate fi interesantă; într-un sistem administrativ, financiar sau medical, aceeași inițiativă poate produce consecințe grave.

Prin urmare, nu aș recomanda nici entuziasmul necondiționat, nici respingerea reflexă. Poziția rezonabilă este evaluarea pe straturi: ce poate face modelul, în ce mediu, cu ce permisiuni, cu ce date, sub ce supraveghere și cu ce mecanism de revenire? Adevărata maturitate nu va fi măsurată prin numărul de agenți lansați, ci prin capacitatea organizației de a controla ceea ce aceștia fac.

Concluzie: nu o explozie a inteligenței, ci începutul industrializării cercetării AI

Nu consider GPT-5.6 o actualizare minoră... nici pe departe. Saltul nu constă doar într-un scor mai mare, ci în reunirea mai multor capabilități: raționare, folosirea instrumentelor, coordonare *multi-agent*, control al calculatorului, evaluare vizuală și participare la experimente de cercetare. Împreună, acestea schimbă rolul modelului din generator de conținut în executant al unui proces.

În același timp, nu aș afirma că am asistat deja la o explozie a inteligenței așa cum vorbeam în articolul „Explozia inteligenței – de la experiment la impact” (Oct. 2025) [21]. Am văzut mai degrabă începutul industrializării cercetării AI: experimente care pot fi formulate, lansate, măsurate și repetate cu o intervenție umană mai redusă. O asemenea accelerare poate comprima ciclurile de dezvoltare și poate amplifica atât progresul, cât și erorile.

Din experiența mea, cele mai promițătoare două direcții sunt *computer use* și *design*-ul orientat spre verificare. Când sistemul poate vedea efectul propriei acțiuni și îl poate compara cu obiectivul, începe să apară o buclă operațională completă. Dar tocmai această buclă trebuie însoțită de reguli clare: înregistrarea acțiunilor, separarea permisiunilor, proveniența surselor, evaluări independente, posibilitatea de *rollback* și atribuirea neechivocă a responsabilității umane. Aceste principii sunt compatibile cu abordarea de management al riscului recomandată de NIST [22].

GPT-5.6 nu elimină autorul, cercetătorul sau profesorul. Îi obligă însă să urce un nivel: de la executarea fiecărui pas la proiectarea procesului, de la producerea materialului la verificarea lui, de la folosirea unui instrument la guvernarea unei mici organizații de agenți. Tocmai aici văd importanța reală a acestei lansări. Nu în promisiunea spectaculoasă că mașina se va îmbunătăți singură, ci în faptul mai sobru și mai imediat că sistemele AI au început să participe efectiv la procesele prin care sunt create, testate și puse în funcțiune următoarele generații de tehnologie.

References

[1] OpenAI, „GPT-5.6: Frontier intelligence that scales with your ambition,” 09 07 2026. [Interactiv]. Available: https://openai.com/index/gpt-5-6/. [Accesat 21 07 2026].

[2] C. Vrabie, „Inteligența artificială după entuziasm: lecții de la AI WEEK Milano 2026,” *All in on Tech (AIoT),* vol. 2, 2026.

[3] OpenAI, „GPT-5.6 System Card,” 09 07 2026. [Interactiv]. Available: https://deploymentsafety.openai.com/gpt-5-6/introduction. [Accesat 21 07 2026].

[4] I. J. Good, „Speculations Concerning the First Ultraintelligent Machine," *Advances in Computers,* vol. 6, pp. 31-88, 1966.

[5] A. Karpathy, „Autoresearch," GitHub, 2026 202. [Interactiv]. Available: https://github.com/karpathy/autoresearch. [Accesat 21 07 2026].

[6] E. Zelikman, E. Lorch, L. Mackey și A. T. Kalai, „Self-Taught Optimizer (STOP): Recursively Self-Improving Code Generation," ArXiv, 03 10 2024. [Interactiv]. Available: https://arxiv.org/abs/2310.02304. [Accesat 21 07 2026].

[7] C. Lu, C. Lu, R. T. Lange, Y. Yamada, S. Hu, J. Foerster, D. Ha și J. Clune, „Towards end-to-end automation of AI research," *Nature,* vol. 651, pp. 914-919, 2026.

[8] Y. Yamada, R. T. Lange, C. Lu, S. Hu, C. Lu, J. Foerster, J. Clune și D. Ha, „The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search," ArXiv, 10 04 2025. [Interactiv]. Available: https://arxiv.org/abs/2504.08066. [Accesat 21 07 2026].

[9] Agent's Last Exam, „Challenge and measure AI agents on economically valuable and real-world tasks.," 2026. [Interactiv]. Available: https://agents-last-exam.org/. [Accesat 21 07 2026].

[10] Artificial Analysis, „Artificial Analysis Coding Agent Benchmarks," 2026. [Interactiv]. Available: https://artificialanalysis.ai/agents/coding-agents. [Accesat 21 07 2026].

[11] Terminal-bench, „benchmarks for ai agents in terminal environments," 2026. [Interactiv]. Available: https://www.tbench.ai/. [Accesat 21 07 2026].

[12] DeepSWE, „Updated execution and grading for the same long-horizon engineering tasks.," 2026. [Interactiv]. Available: https://deepswe.datacurve.ai/blog/deepswe-v1-1. [Accesat 21 07 2026].

[13] BrowseComp, „A benchmark for browsing agents," 2026. [Interactiv]. Available: https://openai.com/index/browsecomp/. [Accesat 21 07 2026].

[14] C. Vrabie, „Civilizații sintetice: cum agenții AI dezvoltă roluri, norme și forme de autonomie socială," *All in on Tech (AIoT),* vol. 2, 2026.

[15] C. Vrabie, „Între Automatizare și Augmentare: Noua Ecuație a Muncii în Era AI," *All in on Tech (AIoT),* vol. 2, 2026.

[16] C. Vrabie, „E-Government 3.0: An AI Model to Use for Enhanced Local Democracies," *Sustainability,* vol. 15, nr. 12, 2023a.

[17] C. Vrabie, „Improving Municipal Responsiveness Through AI-powered Image Analysis in E-Government," *Public Policy and Administration,* vol. 24, nr. 1, pp. 9-23, 2025.

[18] B. Pahonțu, F. Pană-Micu, G. M. Mihăila, L. Movanu și C. Vrabie, „Using Sandboxes for Testing Decisions in the Public Sector," *Administrative Sciences,* vol. 16, nr. 2, 2026.

[19] OpenAI, „ChatGPT Work and Codex," 21 07 2026. [Interactiv]. Available: https://help.openai.com/en/articles/20001275-chatgpt-work-and-codex. [Accesat 21 07 2026].

[20] S. Mavali, D. Pape, J. Evertz, S. Abedini, D. Srivastav, T. Eisenhofer, S. Abdelnabi și L. Schönherr, „No More, No Less: Task Alignment in Terminal Agents," ArXiv, 12 05 2026. [Interactiv]. Available: https://arxiv.org/abs/2605.12233. [Accesat 21 07 2026].

[21] C. Vrabie, „Explozia inteligenței – de la experiment la impact," *All in on Tech (AIoT),* vol. 1, nr. 1, 2025.

[22] National Institute of Standards and Technology, „Artificial Intelligence Risk Management Framework (AI RMF 1.0)," 01 2023. [Interactiv]. Available: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. [Accesat 21 07 2026].

[23] C. Schachtner și C. Vrabie, „University Transfer Architectures for Smart Governance: A Regional Comparison of Scientific Community Building," *Administrative Sciences,* vol. 16, nr. 7, 2026.

[24] Smart-EDU Hub, „SmartY's Adventures," Facultatea de Administrație Publică @ SNSPA, 2026. [Interactiv]. Available: https://www.smart-edu-hub.eu/home/smarty. [Accesat 21 07 2026].

[25] Smart-EDU Hub, „Smart-EDU Hub @ SNSPA YouTube Profile," 2025. [Interactiv]. Available: https://www.youtube.com/@smart-edu-hub. [Accesat 21 07 2026].