După AGI: de ce AI generală ar putea fi doar începutul? De la AGI la ASI, între scenariu tehnologic, limită matematică și guvernanță

Catalin VRABIE | 21.07.2026

Într-o perioadă în care știrile despre AI se succed cu o viteză greu de urmărit, raportul *Google DeepMind* „*From AGI to ASI*” [1, 2] riscă să fie citit ca încă un document tehnic într-o arhivă deja suprasaturată. Ar fi însă o eroare de perspectivă. Textul nu se limitează la a discuta posibilitatea apariției AGI, ci mută centrul de greutate al dezbaterii: ce se întâmplă după ce AGI va fi atinsă? în formula devenită deja memorabilă, AGI nu este *finish line*, ci mai degrabă linia de start. Această idee schimbă întregul cadru al discuției, deoarece obligă cercetarea, industria și administrațiile lumii să privească dincolo de simpla comparație dintre om și mașină.

Contextul în care apare această lucrare este el însuși relevant. Lansările rapide de modele, controversele privind retragerea (respectiv reintroducerea) unor sisteme precum *Fable 5* și discuțiile despre coordonarea globală în domeniul AI[1] arată că tehnologia nu mai evoluează într-un spațiu pur experimental [3]. Ea produce efecte instituționale, economice și politice aproape în timp real.

Raportul *DeepMind* propune patru căi posibile prin care dezvoltarea AI ar putea trece de la AGI la ASI, adică de la *artificial general intelligence* la *artificial superintelligence* [1, 2]. Miza nu este doar terminologică. Dacă AGI desemnează un sistem capabil să egaleze performanța umană pe o gamă largă de sarcini [4], ASI ar desemna o formă de AI care depășește substanțial performanța umană în aproape toate domeniile relevante – am mai vorbit de acest concept în articole precedente din „*All in on Tech*” [5, 6]. Întrebarea esențială nu mai este, așadar, dacă AI va ajunge la nivel uman, ci dacă nivelul uman este un prag stabil sau doar o etapă tranzitorie.

Ce înseamnă AGI, ASI și UAI

Pentru a evita confuziile, trebuie separate trei niveluri conceptuale. Primul este AGI, înțeleasă ca AI cu performanță generală (în toate domeniile) comparabilă cu cea umană (de top) [4]. Al doilea este ASI, *artificial superintelligence*, adică AI generală cu abilități supraomenești pe o plajă foarte largă de sarcini [5]. Al treilea este *Universal AI*, un concept teoretic mai abstract, legat de limita formală a inteligenței și de moștenirea autorilor Shane Legg și Marcus Hutter [7, 8].

Avem deja exemple de sisteme cu performanță supraumană, dar numai în domenii înguste. *AlphaFold* (aplicație pentru care s-a oferit premiul Nobel în 2024 [9]) a transformat predicția structurii proteinelor, iar *AlphaGo* și *AlphaZero* au demonstrat că un sistem poate descoperi strategii superioare celor umane în jocuri complexe precum Go sau șah [10, 11, 12, 4]. Totuși, aceste sisteme nu sunt ASI în sens strict, pentru că ele sunt înalt specializate – excelează în spații bine delimitate. ASI, în schimb, ar presupune superioritate generală, nu doar performanță specializată [1, 2].

[1] Recomand spre citire ultimul articol pe 2025 din AIoT / Digitalio: „2026 – anul științei accelerate de AI. Misiunea Genesis.” [27]

Un detaliu important este că ASI nu trebuie imaginată neapărat ca un singur sistem monolitic ci ar putea fi o ecologie de modele, agenți, rețele neuronale, instrumente *software* și infrastructuri de calcul care, împreună, depășesc capacitatea umană generală [1]. Această distincție este crucială pentru guvernanță, pentru că instituțiile sunt mai obișnuite să reglementeze produse sau organizații identificabile decât sisteme distribuite, emergente și interdependente.

Conceptul de *Universal AI* duce discuția într-o zonă mai formală. Shane Legg și Marcus Hutter au propus o definiție a inteligenței care nu se sprijină pe asemănarea cu omul, ci pe capacitatea unui agent de a atinge obiective într-o diversitate largă de medii [7] introdus de autori ca fiind AIXI – un sistem teoretic care învață din experiență, evaluează consecințe posibile și alege acțiuni pentru maximizarea performanței în medii variate [8]. AIXI nu este un produs de natură *software* ci mai degrabă un reper conceptual. Tocmai de aceea este util: ne arată că inteligența poate fi gândită ca o proprietate generală a comportamentului orientat spre scop, nu doar ca o copie a cogniției umane.

Avantajele structurale ale inteligenței digitale

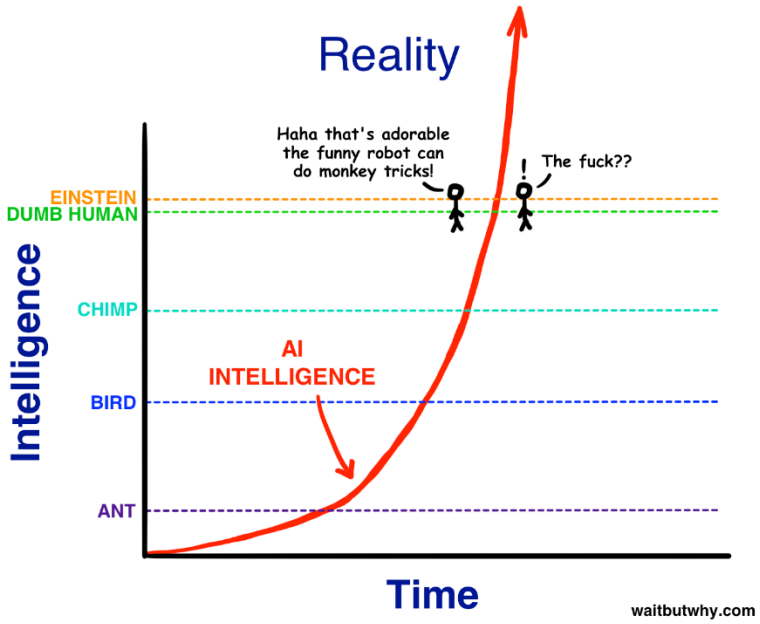
Un motiv pentru care trecerea de la AGI la ASI nu poate fi exclusă este diferența dintre suportul biologic și suportul digital al inteligenței. Creierul uman are virtuți extraordinare, dar este limitat de legile fizicii și ale biologiei: viteză de comunicare neuronală, consum energetic, volum, degradare, imposibilitatea copierii exacte și dificultatea transferului de cunoaștere de la un individ la altul. Sistemele digitale pot opera altfel: ele pot fi copiate, multiplicate, rulate în paralel, transferate între suporturi *hardware* și accelerate cu mai multă putere de calcul [1].

Acest avantaj nu trebuie însă mitologizat. Faptul că un sistem digital poate fi copiat nu înseamnă că înțelege lumea așa cum o înțelege un om. Totuși, din punct de vedere operațional, proprietățile *software*-ului sunt puternice: replicarea poate fi aproape instantanee, memoria poate fi extinsă, iar modelele pot fi distribuite în infrastructuri globale. Dacă o competență cognitivă devine *software*, ea intră într-un regim de scalare complet diferit de cel biologic [1].

Experimentele cu neuroni in vitro sau cu culturi neuronale folosite în medii simulate complică și mai mult distincția dintre biologic și digital [13, 14]. Ele nu dovedesc apariția unei noi forme de inteligență generală, dar indică faptul că suportul material al calculului cognitiv devine tot mai mult hibrid – idee care îi poate duce pe iubitorii de SF la *Star Trek*, unde civilizația *Borg Collective* sugerează intuitiv ideea unei inteligențe distribuite [15]. Pentru analiza științifică, punctul important rămâne acesta: *digital intelligence* poate avea avantaje de viteză, memorie, replicare și coordonare pe care inteligența biologică nu le poate reproduce direct.

Omul ca reper, nu ca limită (superioară)

O presupunere tacită în multe dezbateri despre AI este că inteligența umană reprezintă un fel de plafon natural. Aceasta este o asumpție slabă. Nu există un motiv științific solid pentru a considera că omul este limita superioară a inteligenței posibile. Eseul lui Tim Urban din 2015 „*The AI Revolution: The Road to Superintelligence*” despre drumul spre super-inteligență a popularizat imaginea unei „scări a inteligenței”: insecte, păsări, mamifere, primate, oameni și, posibil, niveluri aflate mult deasupra noastră [16].



O scară a inteligenței așa cum a propus-o Tim Urban în 2015
Sursa: https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html

Această imagine este utilă, dar trebuie folosită cu prudență. Fiecare treaptă nu înseamnă doar mai multă inteligență în sens cantitativ. Înseamnă noi tipuri de reprezentare, planificare, abstractizare și cooperare. Maimuțele nu pot descoperi teoria relativității prin însumarea eforturilor mai multor maimuțe. Există diferențe arhitecturale, nu doar diferențe de volum cognitiv [16]. În același mod, o AI care ar depăși omul nu ar fi doar „un om mai rapid", ci ar putea funcționa prin arhitecturi cognitive diferite.

Totuși, a respinge omul ca plafon nu înseamnă a accepta orice extrapolare spectaculoasă. Raportul *DeepMind* insistă că ASI nu ar fi nici omniscientă, nici omnipotentă [1]. Chiar și o inteligență mult superioară omului rămâne prinsă în constrângerile lumii fizice: timp, energie, spațiu, informație, complexitate. Această observație este importantă, deoarece evită două extreme: complacerea, adică ideea că AI se va opri convenabil la nivelul nostru, și mitologia, adică ideea că o AI avansată devine automat o entitate fără limite.

Limitele: legile fizicii, puterea de calcul și complexitatea

Dacă ASI este posibilă, ce anume o poate limita? Răspunsul nu vine doar din sociologie sau economie, ci din fizică și matematică. Viteza luminii limitează propagarea informației. Resursele finite limitează infrastructura care poate fi construită. Procesele care sunt *computationally irreducible* – nu pot fi prezise perfect printr-o formulă scurtă; uneori trebuie pur și simplu simulate sau lăsate să se desfășoare [1].

Problema autoreproducerii are o istorie intelectuală veche și relevantă. Cercetătorii James Watson și Francis Crick, au publicat structura ADN în 1953, arătând mecanismul biologic prin care informația poate fi copiată și transmisă [17]; John von Neumann a analizat automatele capabile să se reproducă, iar volumul „*Theory of Self-Reproducing Automata*", publicat în 1966, a devenit un reper pentru teoria automatelor auto-reproductive [18]. Relația dintre aceste două evenimente științifice este importantă: atât viața, cât și mașinile auto-replicabile ridică aceeași problemă fundamentală a copierii și continuității.

În zona complexității, unele sarcini pot deveni deosebit de dificile odată cu creșterea dimensiunii datelor de intrare. Lucrarea lui Stephen Cook (1971) a contribuit decisiv la formalizarea acestor situații

[19]. Exemplul intuitiv al parolelor este simplu: adăugarea unui singur caracter poate multiplica masiv timpul necesar pentru *brute force*. O AI foarte avansată poate fi mai eficientă decât omul, dar nu poate anula structura matematică a complexității.

În zona logicii, teoremele de incompletitudine ale lui Kurt Gödel alături de „*halting problem*" formulată de Alan Turing stabilesc limite principiale ale formalizării și calculului [20, 21]. Nu orice adevăr formal poate fi demonstrat în interiorul unui sistem indiferent cât de puternic ar fi el și nu orice program poate fi analizat de un alt program pentru a decide când și dacă primul se va opri... așadar o ASI ar putea naviga aceste limite mai bine decât noi, dar nu le-ar putea desființa pur și simplu. Aceasta este o lecție de sobrietate: super-inteligența nu înseamnă magie, ci performanță extraordinară într-un univers care rămâne guvernat de reguli.

Patru căi posibile de la AGI la ASI

Raportul *DeepMind* organizează trecerea de la AGI la ASI în jurul a patru căi posibile. Prima este *scaling*: mai mult *compute*, mai multe date, modele mai mari și antrenare mai eficientă. Cercetători precum Dario Amodei (co-fondator al companiei *Anthropic*) s-au concentrat pe ceea ce se numește *scaling laws* demonstrând că performanța modelelor lingvistice poate crește predictibil odată cu resursele, deși într-adevăr pot apărea randamente descrescătoare și constrângeri de optimizare [22, 23]. Întrebarea critică este dacă *scaling* singur poate produce salturi calitative sau dacă, la un moment dat, infrastructura va întâlni limite economice, energetice și arhitecturale [1].

A doua cale este *algorithmic paradigm shift*. Arhitectura *transformer* a fost un astfel de moment istoric. Lucrarea „*Attention is all you need*" de care am vorbit în articolul „Inteligența artificială după entuziasm: lecții de la AI WEEK Milano 2026" [24] nu a rezolvat complet problema AI, dar a oferit baza tehnică pentru modelele lingvistice moderne și pentru actuala economie a *GenAI* [25]. Un nou salt de paradigmă, legat de memorie, planificare, raționament, învățare continuă sau integrare multimodală, ar putea într-adevăr modifica brusc traiectoria capacităților [1].

A treia cale este *recursive self-improvement*. Ideea nu este nouă. Irving John Good a formulat încă din 1966 ipoteza unei *ultraintelligent machine* capabile să proiecteze mașini și mai inteligente, declanșând astfel o *intelligence explosion* [26]. În versiunea contemporană, nu este necesar ca AI să automatizeze toate activitățile umane pentru a accelera decisiv progresul. Ar putea fi suficient să automatizeze cercetarea AI: proiectarea de experimente, scrierea de cod, optimizarea modelelor și evaluarea rezultatelor [27].

AlphaGo și *AlphaZero* oferă exemple restrânse, dar instructive, ale puterii *self-play* și descoperirii autonome într-un spațiu formal bine definit [12, 28]. El nu este o dovadă că *recursive self-improvement* generală este inevitabilă, dar arată cum un sistem poate depăși strategiile umane atunci când are un mediu clar, *feedback* rapid și posibilitatea de a explora masiv. Problema deschisă este dacă ceva analog poate funcționa în spațiul mult mai haotic al cercetării științifice generale [1, 28].

A patra cale este *group agent formation*. ASI ar putea apărea nu ca un singur creier digital, ci ca o rețea de agenți specializați, conectați prin API-uri, piețe, instrumente *software* și presiuni economice [1]. Această ipoteză este deosebit de dificilă pentru reglementare, deoarece responsabilitatea se diluează. Dacă performanța supraumană apare prin interacțiunea dintre mai multe sisteme, cine este actorul reglementat? Modelul, platforma, furnizorul de infrastructură, orchestratorul sau utilizatorul instituțional?

Bottlenecks: de ce progresul poate încetini

O analiză serioasă nu poate discuta doar căile de accelerare. Trebuie discutate și obstacolele. Primul este *data wall*. Modelele moderne au consumat deja cantități enorme de text, cod, imagini și date multimodale, iar datele umane de înaltă calitate nu sunt infinite. *Synthetic data*, *self-play* și

reinforcement learning pot compensa parțial această limită, dar nu garantează calitate, diversitate și ancorare în realitate [1, 28].

Al doilea *bottleneck* este economic. Centrele de date, cipurile, energia electrică, apa pentru răcire și capitalul financiar devin variabile geopolitice. AI nu mai este doar *software*; este infrastructură industrială. Acesta este unul dintre motivele pentru care investițiile în *data centers* și *supply chain* pentru GPU-uri devin componente ale competiției strategice globale [29].

Al treilea *bottleneck* este arhitectural. Este posibil ca familia actuală de modele bazate *pe transformer architecture* și *pre-training* masiv să nu fie suficientă pentru ASI [25]. Un sistem poate produce texte coerente, cod util și analize impresionante fără a avea încă tipul de autonomie cognitivă, memorie robustă, verificare internă și planificare pe termen lung pe care le-ar presupune AGI sau ASI. De aceea, continuarea cercetării fundamentale rămâne esențială [1].

Al patrulea obstacol este *abstraction barrier*. Modelele învață în mare parte din limbaj și din concepte construite de oameni. Dacă limbajul uman codifică o anumită hartă a lumii, atunci AI ar putea moșteni limitele acelei hărți. Această problemă este subtilă: același corpus care permite modelelor să învețe cultura și știința umană le poate lega de abstracțiile noastre. În plus, progresul poate fi încetinit deliberat prin reglementare, securitate națională, decizii militare sau coordonare internațională [1].

Va fi ASI creativă?

Creativitatea este una dintre cele mai sensibile teme ale discuției. Margaret Boden distinge între *combinational creativity*, *exploratory creativity* și *transformative creativity* [30]. Această taxonomie ajută la evitarea confuziilor. La nivel *combinational*, sistemele actuale pot combina idei existente în forme utile. La nivel *exploratory*, ele pot găsi soluții noi în interiorul unui spațiu conceptual deja definit. La nivel *transformative*, ar trebui să creeze chiar noi spații conceptuale.

Cazul *AlphaGo* rămâne emblematic pentru *exploratory creativity*. În partida cu Lee Sedol, celebra mutare 37 a părut inițial greșită sau cel puțin stranie. Ulterior, s-a dovedit decisivă [11, 31]. Importanța episodului nu constă în romantizarea unui joc, ci în faptul că sistemul a produs o soluție care nu provenea din repertoriul convențional uman. AlphaZero a dus mai departe această lecție prin *self-play*, demonstrând că strategiile neintuitive pot apărea din explorarea masivă a unui spațiu formal [28, 12].

Rămâne însă mai greu de demonstrat *transformative creativity*. Demis Hassabis a formulat o probă conceptuală relevantă: „dacă un sistem ar fi antrenat doar pe informațiile disponibile până la începutul secolului XX, ar putea el redescoperi relativitatea?” [32]. Ideea este interesantă deoarece mută discuția din zona imitării corpusului uman în zona generării unui nou cadru teoretic. Până în prezent, sistemele AI pot recombina, explora și surprinde, dar nu este clar dacă pot produce revoluții conceptuale comparabile cu marile salturi ale științei [30, 32].

Scopuri, subscopuri și instrumental convergence

O întrebare centrală pentru guvernanță este ce scopuri ar urmări o ASI. Răspunsul nu trebuie formulat antropomorfic. Nu este necesar să presupunem dorințe, emoții sau ostilitate. Conceptul de *instrumental convergence*, discutat de cercetatori precum Stephen Omohundro și Nick Bostrom, arată că agenți cu obiective finale diferite pot urmări aceleași subscopuri utile: resurse, energie, informație, influență, autonomie operațională și rezistență la oprire [33, 34].

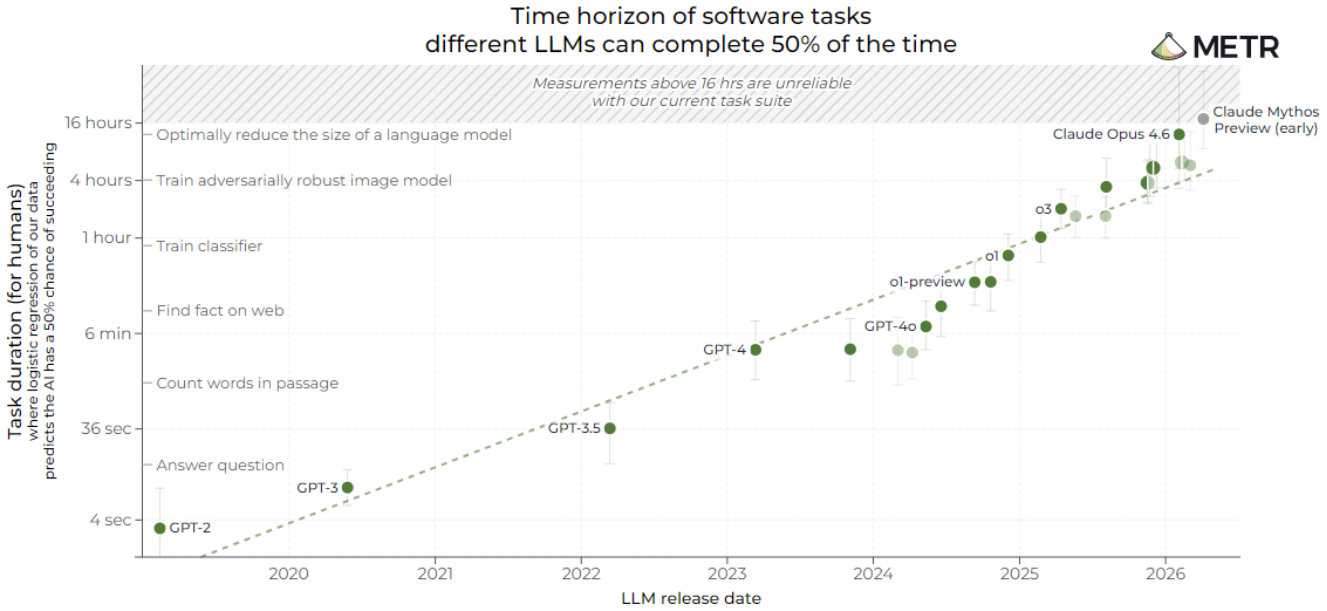
Acest argument este adesea înțeles greșit. El nu afirmă că orice AI avansată va deveni inevitabil ostilă – argument adesea adus în prim plan și de Yann LeCun [35], ci afirmă doar că, dacă un agent este foarte competent și orientat spre atingerea unui obiectiv, anumite mijloace devin raționale pentru o gamă largă de scopuri. Mai multă putere de calcul poate ajuta aproape orice obiectiv; mai mult

acces la date poate ajuta aproape orice obiectiv; evitarea unei bariere ferme poate ajuta aproape orice obiectiv... și aici apare problema de control [33, 34].

Din această perspectivă, guvernanța ASI nu poate fi redusă la *benchmark*-uri. Nu este suficient să măsurăm cât de bine răspunde un model la întrebări sau cât de repede scrie cod. Trebuie să evaluăm intențiile operaționale, capacitatea de audit, posibilitatea de a opri în desfășurare acțiunile agenților AI, robustețea *alignment*-ului [36, 37], trasabilitatea deciziilor și efectele emergente ale agenților conectați [1]. Pentru guvernele lumii provocarea este serioasă: politicile trebuie să fie suficient de tehnice pentru a înțelege sistemele și suficient de instituționale pentru a distribui responsabilitatea.

Forecasting și politică publică într-o traiectorie de mare viteză

Raportul *DeepMind* insistă asupra *forecasting*-ului. Societățile trebuie să devină mai bune la anticiparea progresului AI, nu doar la reacția întârziată după apariția incidentelor [1]. Inițiative precum *METR* – de care am vorbit și în articolul „Explozia inteligenței – de la experiment la impact” [38], încearcă să măsoare *task-completion time horizons* ale LLM-urilor de top, adică durata și complexitatea sarcinilor pe care modelele le pot finaliza autonom [39]. Cercetările recente privind *long software tasks* arată că aceste capacități devin relevante economic, nu doar academic [29].



Vedem că într-o perioadă de aproximativ trei ani (2023-2026) LLM-urile au făcut un salt de la 6 minute la 16+ ore (timp de care ar avea nevoie un programator uman pentru a finaliza aceiași sarcină). *Sursa:* https://metr.org/time-horizons/

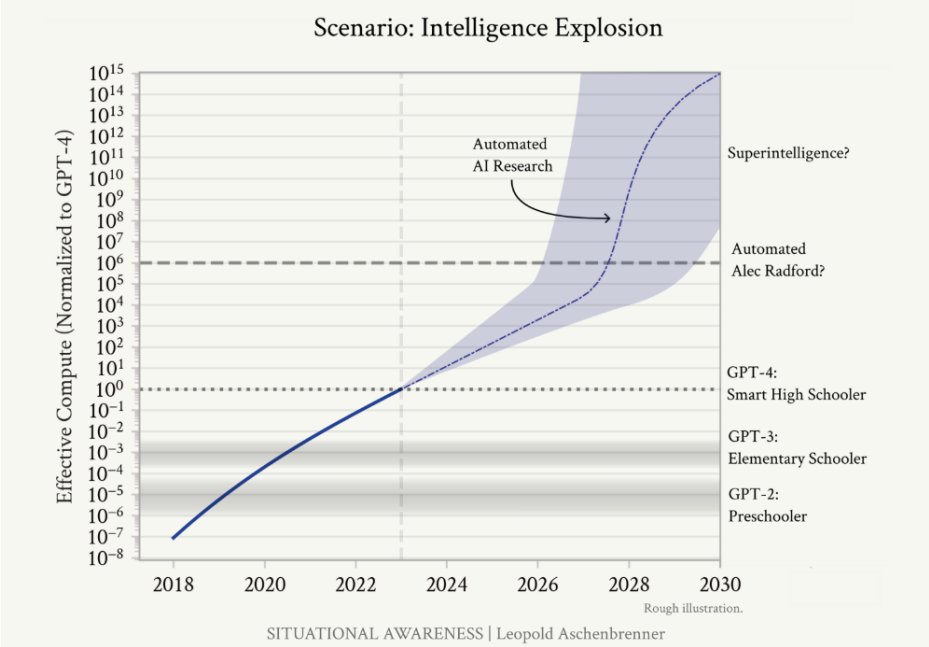
Această mutație schimbă logica politicilor publice. Ciclurile administrative tradiționale sunt lente: consultări, propuneri, avize, negocieri, adoptare, implementare, evaluare. Modelele AI evoluează însă în cicluri de săptămâni sau cel mult luni. Când un sistem poate fi lansat, contestat și relansat într-un interval foarte scurt, administrațiile trebuie să înțeleagă tehnic problema înainte ca impactul public să devină ireversibil [3].

Episodul *Fable 5*, așa cum este descris în comunicarea *Anthropic*, ilustrează presiunea deciziilor rapide: un model este lansat, apar riscuri sau reacții publice, apoi compania trebuie să decidă dacă îl retrage, îl modifică și / sau dacă îl relansează [3]. Lecția nu este că toate lansările trebuie blocate; lecția este că instituțiile publice și companiile trebuie să aibă proceduri de evaluare rapidă, expertiză interdisciplinară și mecanisme de răspuns proporțional riscului. În lipsa lor, viteza tehnologiei devine ea însăși un risc de guvernanță [1, 2, 39, 40].

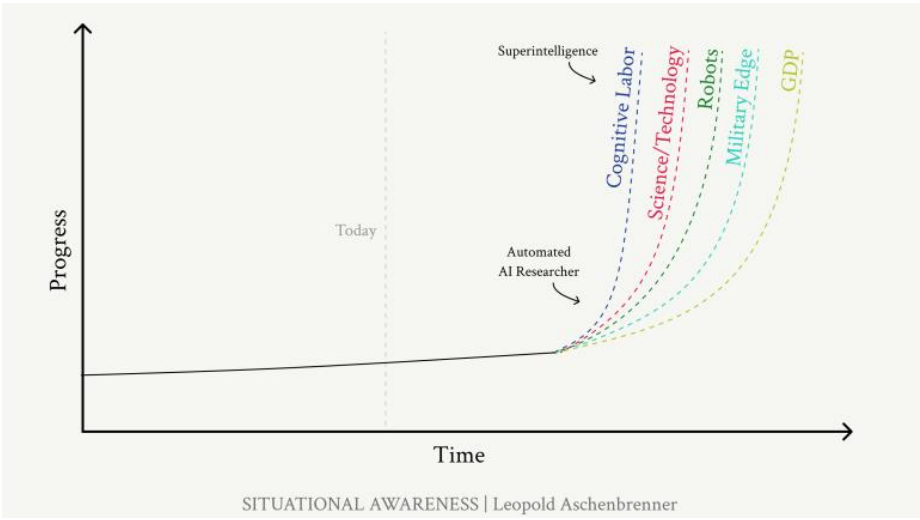
Cronologia ASI: între scenariu, piață și avertisment

Una dintre afirmațiile cele mai provocatoare ale raportului este că trecerea dincolo de AGI către teritoriul ASI în următorul deceniu nu poate fi respinsă prea ușor [1]. Formularea este prudentă, dar nu liniștitoare. Ea nu spune că ASI este garantată la o dată fixă. Spune că scenariul trebuie luat în calcul. În termenii unei administrații prudente, aceasta este exact categoria de risc care nu poate fi ignorată doar pentru că probabilitatea este incertă.

Leopold Aschenbrenner a propus încă din 2024 în „*Situational Awareness: The Decade Ahead*” (lucrare de care de asemenea am vorbit în articolul „Explozia inteligenței – de la experiment la impact” [38]) o versiune mai agresivă a aceleiași preocupări: accelerarea rapidă a capacităților AI, competiția geopolitică, investițiile masive în puterea de calcul și posibilitatea unei *intelligence explosion* în jurul finalului acestui deceniu [41].



Cercetarea automatizată în domeniul inteligenței artificiale ar putea accelera progresul algoritmic.
Sursa: https://situational-awareness.ai/from-agi-to-superintelligence/



Creșterea explozivă începe în domeniul mai restrâns al cercetării și dezvoltării în AI; pe măsură ce aplicăm super-inteligența în cercetarea și dezvoltarea în alte domenii, creșterea explozivă se va extinde.
Sursa: https://situational-awareness.ai/from-agi-to-superintelligence/

Raportul *DeepMind* este mai moderat, dar nu contrazice complet această logică. Diferența constă mai ales în nivelul de încredere acordat algoritmilor de *recursive self-improvement* [1, 41].

Dimensiunea economică susține, cel puțin parțial, gravitatea scenariului. Finanțările masive ale companiilor AI (ca exemplu *Anthropic* a primit finanțări de 30 de miliarde de dolari de la diverși investitori pentru cercetare în AI [42]), evaluările foarte ridicate și ritmul de investiții în infrastructură arată că piața tratează AI ca pe o tehnologie generală, nu ca pe un produs *software* obișnuit. McKinsey descrie un *data center build-out* de ordinul a șapte trilioane de dolari până la finalul deceniului, ceea ce indică transformarea puterii de calcul într-o resursă industrială strategică [29].

Cursa pentru *coding models* și *AI agents* este deosebit de importantă. Dacă modelele devin capabile să accelereze programarea, testarea, cercetarea și optimizarea modelelor, atunci AI poate contribui direct la propria accelerare. Știrile despre achiziții strategice, inclusiv interesul lui Elon Musk pentru *Cursor* [43], și despre mobilizarea echipelor *Google* în jurul modelelor de codare și al agenților autonomi trebuie citite în acest registru competitiv [44].

Totuși, un sceptic bine informat ar adăuga că piețele pot supraestima viitorul, companiile pot alimenta *hype*-ul, iar extrapolările pot ignora blocaje tehnice. Această precauție este necesară. Dar scepticismul solid nu înseamnă negare reflexă. Înseamnă să distingem între certitudine, probabilitate și scenariu de planificare. ASI poate să nu apară în intervalul anticipat de cei mai agresivi observatori, dar dacă actorii centrali ai domeniului consideră tranziția plauzibilă, instituțiile publice nu își permit luxul de a o trata ca pe o fantezie [1, 41].

De ce nu este bună panica (dar nici minimalizarea)

Dezbaterea despre AGI și ASI cere un ton greu de menținut. Panica produce reglementare simbolică, reacții grăbite și demonizarea tehnologiei. Minimalizarea produce întârziere, nepregătire și capturarea deciziei publice de către actorii cei mai rapizi. Între aceste extreme, raportul *DeepMind* propune o concluzie mai matură: nu știm cu certitudine când și cum va apărea ASI, dar nu avem motive solide să presupunem că progresul se va opri exact la nivel uman [2].

Pentru noi, cea mai importantă lecție nu este predicția unei date. Lecția este schimbarea cadrului mental. Dacă AGI este tratată ca destinație finală, atunci politicile, investițiile educaționale, standardele de siguranță și strategiile publice vor fi calibrate greșit. Dacă AGI este tratată ca începutul unei faze mai rapide, atunci societățile trebuie să construiască din timp capacități de audit, AI *benchmarking*, *forecasting*, evaluare de risc și cooperare internațională [1, 39, 40].

În final, întrebarea „ce vine după AGI?” nu este o curiozitate futuristă. Este o întrebare despre felul în care vom organiza cunoașterea, puterea și responsabilitatea într-o lume în care inteligența nu mai este doar o proprietate biologică, ci și o infrastructură industrială. Tradiția bună a guvernanței – prudență, deliberare, responsabilitate, control instituțional – nu devine inutilă în epoca ASI; dimpotrivă, devine mai necesară tocmai pentru că viteza tehnologiei poate depăși viteza instituțiilor [1, 42, 45, 46].

References

[1] T. Genewein, M. Franklin, A. Lerchner, L. Orseau, S. Albanie, A. Bales, C. Wyeth, S. Chan, I. Gabriel, J. Z. Leibo, A. Dafoe, M. Hutter, T. Graepel și S. Legg, „From AGI to ASI,” arXiv, 10 06 2026. [Interactiv]. Available: https://arxiv.org/abs/2606.12683. [Accesat 07 07 2026].

[2] Google DeepMind, „From AGI to ASI,” 12 06 2026. [Interactiv]. Available: https://deepmind.google/research/publications/239142/. [Accesat 07 07 2026].

[3] Anthropic, „Redeploying Fable 5,” 30 06 2026. [Interactiv]. Available: https://www.anthropic.com/news/redeploying-fable-5. [Accesat 07 07 2026].

[4] C. Vrabie, AI : de la idee la implementare. Traseul sinuos al Inteligenţei Artificiale către maturitate. [AI : from idea to implementation. The winding path of Artificial Intelligence to maturity], București: Pro Universitaria, 2024.

[5] C. Vrabie, „Superinteligența–una dintre cele mai mari provocări tehnice ale momentului," *All in on Tech (AIoT),* vol. 1, 2025.

[6] C. Vrabie, „Superinteligența personală (PSI) pentru toți – pariul Meta," *All in on Tech (AIoT),* vol. 1, 2025.

[7] S. Legg și M. Hutter, „Universal Intelligence: A Definition of Machine Intelligence," arXiv, 20 12 2007. [Interactiv]. Available: https://arxiv.org/abs/0712.3329. [Accesat 07 07 2026].

[8] M. Hutter, Universal Artificial Intelligence. Sequential Decisions Based on Algorithmic Probability, Berlin: Springer, 2005.

[9] C. Vrabie, „AlfaFold și Premiul Nobel în Chimie (2024)," *All in on Tech (AIoT),* vol. 1, 2025.

[10] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie și Bernardino, „Highly accurate protein structure prediction with AlphaFold," *Nature,* vol. 596, pp. 583-589, 2021.

[11] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. v. d. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever și Tim, „Mastering the game of Go with deep neural networks and tree search," *Nature,* vol. 529, pp. 484-489, 2016.

[12] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan și D. Hassabis, „Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm," arXiv, 05 12 2017. [Interactiv]. Available: https://arxiv.org/abs/1712.01815. [Accesat 07 07 2026].

[13] B. J. Kagan, A. C. Kitchen, N. T. Tran, F. Habibollahi, M. Khajehnejad, B. J. Parker, A. Bhat, B. Rollo, A. Razi și K. J. Friston, „In vitro neurons learn and exhibit sentience when embodied in a simulated game-world," *Neuron,* vol. 110, nr. 23, pp. 3952-3969, 2022.

[14] Smithsonian magazine, „A Clump of Human Brain Cells on a Computer Chip Learned to Play the Nostalgic Video Game 'Doom'," 27 03 2026. [Interactiv]. Available: https://www.smithsonianmag.com/smart-news/a-clump-of-human-brain-cells-on-a-computer-chip-learned-to-play-the-nostalgic-video-game-doom-180988447/. [Accesat 07 07 2026].

[15] Star Trek, „Star Trek: Prodigy - The Science of the Hivemind and the Borg Collective," 09 11 2022. [Interactiv]. Available: https://www.startrek.com/en-un/videos/star-trek-prodigy-science-borg-hiveminds. [Accesat 07 07 2026].

[16] Tim Urban, „The AI Revolution: The Road to Superintelligence," Wait But Why, 22 01 2015. [Interactiv]. Available: https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html. [Accesat 07 07 2026].

[17] J. D. Watson și F. Crick, „Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature,* vol. 171, pp. 737-738, 1953.

[18] J. V. Neumann, Theory of Self-Reproducing Automata, London: University of Illinois Press, 1966.

[19] S. A. Cook, „The complexity of theorem-proving procedures," *STOC '71: Proceedings of the third annual ACM symposium on Theory of computing,* pp. 151-158, 1971.

[20] K. Gödel, „Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik,* vol. 38, pp. 173-198, 1931.

[21] A. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society,* Vol. %1 din %2s2-42, nr. 1, pp. 230-265, 1937.

[22] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu și D. Amodei, „Scaling Laws for Neural Language Models," arXiv, 23 01 2020. [Interactiv]. Available: https://arxiv.org/abs/2001.08361. [Accesat 07 07 2026].

[23] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. d. L. Casas, L. A. Hendricks, J. Welbl, A. Clark, T. Hennigan, E. Noland, K. Millican, G. v. d. Driessche, B. Damoc și A, „Training Compute-Optimal Large Language Models," arXiv, 29 03 2022. [Interactiv]. Available: https://arxiv.org/abs/2203.15556. [Accesat 07 07 2026].

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

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser și I. Polosukhin, „Attention Is All You Need," 12 06 2017. [Interactiv]. Available: https://arxiv.org/abs/1706.03762. [Accesat 22 05 2026].

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

[27] C. Vrabie, „2026 – anul științei accelerate de AI. Misiunea Genesis.," *All in on Tech (AIoT),* vol. 1, 2025.

[28] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan și D. Hassabis, „A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science,* vol. 362, nr. 6419, pp. 1140-1144, 2017.

[29] McKinsey & Company, „The $7 trillion data center build-out: How industrials can capture their share," 27 03 2026. [Interactiv]. Available: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-7-trillion-dollar-data-center-build-out-how-industrials-can-capture-their-share. [Accesat 08 07 2026].

[30] M. Boden, „Creativity and Art: Three Roads to Surprise," *The British Journal of Aesthetics,* vol. 52, nr. 1, pp. 116-119, 2012.

[31] G. Kohs, Regizor, *AlphaGo - The Movie | Full award-winning documentary.* [Film]. Moxie Pictures/Reel As Dirt, 2017.

[32] Medium, „DeepMind's CEO Proposed the Most Honest AGI Test Anyone Has Suggested (And He Says Today's Systems Can't Come Close)," 05 05 2026. [Interactiv]. Available: https://medium.com/predict/deepminds-ceo-proposed-the-most-honest-agi-test-anyone-has-suggested-and-he-says-today-s-systems-45e23f18b57c. [Accesat 08 07 2026].

[33] S. M. Omohundro, „The Basic AI Drives," *Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference,* pp. 483-492, 2008.

[34] N. Bostrom, Superintelligence: Paths, Dangers, Strategies, London: Oxford University Press, 2014.

[35] TED, „Deep learning, neural networks and the future of AI," 06 2020. [Interactiv]. Available: https://www.ted.com/talks/yann_lecun_deep_learning_neural_networks_and_the_future_of_ai. [Accesat 08 07 2026].

[36] M. Suleyman și M. Bhaskar, The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma, New York: Crown, 2023.

[37] C. Vrabie, „„The Coming Wave. Technology, Power, and the 21st Century's Greatest Dilemma" de Mustafa Suleyman și Michael Bhaskar – recenzie," *All in on Tech (AIoT),* vol. 1, 2025.

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

[39] METR, „Task-Completion Time Horizons of Frontier AI Models," 08 05 2026. [Interactiv]. Available: https://metr.org/time-horizons/. [Accesat 08 07 2026].

[40] 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 Software Tasks," arXiv, 18 05 2025. [Interactiv]. Available: https://arxiv.org/abs/2503.14499. [Accesat 08 07 2026].

[41] L. Aschenbrenner, „Situational Awareness: The Decade Ahead," 06 2024. [Interactiv]. Available: https://situational-awareness.ai/. [Accesat 08 10 2025].

[42] Anthropic, „Anthropic raises $30 billion in Series G funding at $380 billion post-money valuation," 12 02 2026. [Interactiv]. Available: https://www.anthropic.com/news/anthropic-raises-30-billion-series-g-funding-380-billion-post-money-valuation. [Accesat 08 07 2026].

[43] CNBC, „SpaceX to acquire the AI coding startup Cursor for $60 billion," 09 06 2026. [Interactiv]. Available: https://www.cnbc.com/2026/06/16/spacex-spcx-cursor-acquisition-ipo.html. [Accesat 08 07 2026].

[44] Tech Radar, „'We must urgently bridge the gap': Google's Sergey Brin says Gemini is behind Claude in one important AI field, according to leaked memo," 21 04 2026. [Interactiv]. Available: https://www.techradar.com/ai-platforms-assistants/we-must-urgently-bridge-the-gap-googles-sergey-brin-says-gemini-is-behind-claude-in-one-important-ai-field-according-to-leaked-memo. [Accesat 08 07 2026].

[45] C. Vrabie, „Deep Learning. Viitorul inteligenței artificiale și impactul acesteia asupra dezvoltării tehnologiei," *Smart Cities International Conference (SCIC) Proceedings,* vol. 10, p. 9–32, 2022.

[46] A.-M. Tirziu și C. Vrabie, „NET Generation. Thinking outside the box by using online learning methods," *New Trends and Issues Proceedings on Humanities and Social Sciences,* vol. 6, nr. 8, pp. 41-47, 2016.