„Firul roșu” al inteligenței artificiale: scandalul *Anthropic*–Pentagon între etică, lege și logica războiului

Catalin VRABIE | 01.03.2026

În ultimele zile ale lunii februarie 2026, o dispută care mocnea de luni bune între compania de inteligență artificială *Anthropic* (dezvoltatoarea modelului AI numit *Claude* [1, 2] pe care l-am amintit în mai multe articole din *All in on Tech* [3, 4, 5]) și aparatul american de apărare a ieșit la suprafață cu forța unui război instituțional [6, 7, 8]. Potrivit relatărilor presei internaționale, administrația de la *Washington* a decis să oprească utilizarea tehnologiei *Anthropic* în toate agențiile federale, iar în paralel Pentagonul a încadrat compania într-o categorie de „risc” (*supply-chain risk* mai exact), semnal cu impact direct asupra contractorilor și integratorilor care folosesc modelele *Claude* în infrastructuri guvernamentale [9, 10, 11].

Sursa: https://www.cnbc.com/2026/02/27/trump-anthropic-ai-pentagon.html

Dincolo de titlurile sonore și de reflexul inevitabil al polarizării politice, conflictul atinge o întrebare veche, aproape clasică, a modernității tehnice: cine stabilește limitele utilizării unei tehnologii cu potențial strategic – producătorul, statul sau legea [12, 13]? În forma sa cea mai dură, întrebarea se transformă în: poate un actor privat să traseze un „fir roșu” într-un domeniu în care statul pretinde, tradițional, că deține ultima instanță a deciziei [14, 15, 16]?

În articole apărute în presa internațională, poziția *Anthropic* se sprijină pe două poziții de principiu: (1) refuzul de a permite folosirea modelelor sale pentru arme complet autonome fără control uman și (2) refuzul de a facilita supravegherea domestică a cetățenilor (aici americani dar cred că putem extrapola cu ușurință la noi toți) [17, 18, 8]. În contrapondere,

poziția Pentagonului apare prezentată ca o cerere de acces fără restricții, formulată în termenii „oricărei utilizări legale”, sub argumentul că decizia aparține autorității publice, nu ai unei companii private prin *Terms and conditions* [19, 20, 9, 21].

Acest articol își propune să urmărească presa internațională importantă care s-a aplecat pe subiect păstrând disciplina analitică specifică mediului academic cu mențiunea că în astfel de crize, faptele, interpretările și retorica se amestecă. De aceea, nu este suficient să redăm doar narațiunea uneia dintre părți ci trebuie să privim scandalul apărut ca pe un nod unde se întâlnesc: (1) politica de siguranță a unui laborator AI de top cu (2) practica instituțională a achizițiilor militare, (3) tensiunile culturale dintre *Silicon Valley* și statul „securitar” [22, 23], și (4) o problemă tehnică adesea ignorată în *talk-show*-uri: LLM-urile sunt sisteme probabilistice în care erorile (rare, dar severe) apărute la marginea distribuției datelor, contează enorm în contexte de viață și moarte [9, 24].

De la promisiunea *safety-first* la ciocnirea cu statul

Anthropic s-a construit public în jurul unei identități: aceea de laborator AI care pune siguranța dezvoltării înaintea vitezei [25, 26]. În relatările apărute în presa internațională încă de la înființarea companiei, se insistă asupra faptului că o parte dintre fondatorii ei au plecat dintr-un laborator rival (este vorba de *OpenAI* [27, 28, 29]) din dezacord privind ritmul lansărilor și seriozitatea măsurilor de control al riscurilor. Indiferent de simpatiile pe care le poate trezi o asemenea poveste, ea a funcționat ca un „contract moral” cu publicul: promisiunea că dezvoltarea modelelor de top nu va fi lăsată exclusiv în grija concurenței și a pieței, ci va fi supusă unei guvernanțe interne [30, 31].

Instrumentul principal al acestei guvernanțe a fost un cadru de politici cunoscut sub numele de *Responsible Scaling Policy* (RSP). Pe scurt, RSP este o încercare de a lega deciziile de dezvoltare (antrenare, scalare, lansare) de praguri de risc și de un set de obligații: evaluări, auditări, măsuri de securitate și limitări de utilizare. Din perspectiva tradiției instituționale, RSP seamănă cu un „cod tehnic de conduită”, o versiune corporativă a ideii că puterea (aici: puterea modelului) cere responsabilitate [30, 32, 33].

Problema este că, odată ce un sistem devine suficient de util pentru administrația publică și, mai ales, pentru apărare, „responsabilitatea” nu mai este o dezbatere abstractă; ea se lovește de realități precum: operațiuni clasificate, urgență, presiune geopolitică și o logică instituțională în care „a nu avea instrumentul” poate fi perceput ca un risc strategic în sine.

Într-o lume ordonată, limitele sunt stabilite prin lege și control democratic; în lumea reală însă, limitele sunt adesea negociate prin contracte, standarde și etichete birocratice… iar această negociere a devenit miezul scandalului.

Motivul declanșator a fost utilizarea militară și „problema intermediarului”

În reconstrucțiile vehiculate de presa internațională, ruptura ar fi fost precipitată de suspiciunea că un model *Anthropic* a ajuns să fie folosit, direct sau indirect, în contexte operaționale sensibile, prin intermediul unor platforme de integrare și analiză[1]. Aici apare un

[1] Este vorba de capturarea președintelui Venezuelei, Nicolás Maduro. *Axios* [61] și *Wall Street Journal* [62, 63] au relatat ulterior desfășurării evenimentelor că Pentagonul a folosit modelul *Claude* „în timpul operațiunii” și nu doar la pregătirea operațiunii. Mai departe, *Reuters* a preluat informația atribuită WSJ și a subliniat că nu a putut verifica independent relatarea și că *Departament of Defence*, Casa Albă, *Anthropic* și *Palantir* nu au răspuns imediat solicitărilor; *Reuters* notează și că politicile *Anthropic*

detaliu tehnic esențial și adesea ignorat: în ecosistemul modern al AI, laboratorul care dezvoltă modelul nu controlează întotdeauna implementarea finală [34, 35].

De regulă, există un *pipeline*: furnizorul de model → furnizorul de infrastructură (cloud, rețea, *runtime*) → integratorul (contractorul) → aplicația finală → utilizatorii operaționali [36]. În mediile militare și guvernamentale, aceasta include și elemente suplimentare precum: compartimentare, sisteme clasificate, mecanisme de logare și audit cu acces restricționat, și o arhitectură contractuală în care responsabilitățile sunt distribuite [37, 38, 39]. Într-o asemenea situație, o companie ca *Anthropic* poate controla: ce oferă, cum licențiază, ce interdicții include în termeni, dar nu poate controla complet în ce fel un contractor „împachetează" modelul, cum îl conectează la surse de date, sau cum îl folosește un analist într-un scenariu de urgență [40].

Din punct de vedere instituțional, aici se nasc două neînțelegeri tipice:

- Compania spune: „nu am autorizat această utilizare și avem interdicții clare".
- Statul spune: „dacă ai lăsat tehnologia să ajungă în ecosistem, nu poți pretinde că dictezi ce se întâmplă după aceea, mai ales când vorbim de securitate națională".

Dacă această descriere este corectă, scandalul nu este doar moral; este și arhitectural sau de *design*. Arată că „firul roșu" declarat de un laborator poate fi ocolit prin designul *pipeline*-ului. Iar când un fir roșu poate fi ocolit, el încetează să fie „fir" și devine doar „preferință".

În astfel de cazuri, reacția instituțiilor de apărare este previzibilă: dacă furnizorul ridică obiecții, el poate fi încadrat drept „partener instabil". În limbajul achizițiilor, „instabilitatea" se traduce în risc: risc de continuitate, risc de conformitate, risc de interoperabilitate etc., așa se ajunge la etichete precum *supply-chain risk*, care sunt mai mult decât stigmate... sunt instrumente de guvernanță [9, 20, 41].

Miezul disputei: „orice utilizare legală" versus „firul roșu"

În orice stat modern, există un principiu de bază: decizia privind folosirea forței aparține instituțiilor legitime ale statului, în cadrul legii [42, 43]. Aceasta este ideea care stă în spatele argumentului „oricărei utilizări legale"... numai că, în domeniul AI, legalitatea nu acoperă tot spectrul riscului, din două motive:

- Opacitatea: multe utilizări relevante pentru securitate sunt clasificate, iar auditul public este limitat [44, 45].
- Noutatea: legea tinde să urmărească tehnologia, nu să o preceadă. În perioade de tranziție tehnică, există zone gri în care „legal" poate însemna doar „încă nereglementat" [46, 47].

De aceea, companiile încearcă să introducă restricții contractuale sau tehnice. *Anthropic* ar fi insistat, potrivit presei internaționale, pe două linii: (a) fără arme complet autonome care selectează și angajează ținte fără control uman semnificativ; (b) fără supraveghere domestică [19, 20].

Un sceptic bine informat ar ridica imediat întrebarea: de ce tocmai aceste două? Răspunsul este relativ clasic, ba poate chiar tradițional:

interzic folosirea *Claude* pentru a susține violența, a proiecta arme sau a realiza supraveghere [64], iar WSJ a indicat că implementarea ar fi venit prin parteneriatul *Anthropic–Palantir*.

- În cazul armelor autonome, problema este responsabilitatea: cine răspunde pentru o decizie letală luată de un sistem? În tradiția dreptului războiului și a disciplinei militare, există legături clare de comandă și răspundere. Automatizarea completă riscă să dilueze răspunderea într-o mare de „nu eu”, ci dezvoltatorul, integratorul, operatorul, comandantul etc.
- În cazul supravegherii domestice, problema este dreptul constituțional și tradiția liberală: statul de drept se protejează tocmai prin limite puse puterii executive, iar tehnologiile de supraveghere sunt, istoric, tentante și abuzabile [22].

Pe de altă parte, un alt sceptic – de data aceasta din interiorul aparatului de securitate – ar formula obiecția: dacă statul acceptă ca furnizorii privați să impună limite, atunci suveranitatea operațională este fragmentată. Astăzi un laborator refuză autonomia, mâine altul refuză un anumit teatru de operațiuni, poimâine un altul blochează o funcție critică. În logica apărării, o astfel de dependență de „preferințele morale” ale unor consilii corporative este percepută ca vulnerabilitate strategică.

Aici se vede că disputa nu este doar despre „bine” și „rău”, ci despre cine are dreptul să fixeze normele într-o tehnologie nouă, fără precedent stabil.

Instrumentele de presiune: etichete, contracte și „lanțul de aprovizionare”

Potrivit relatărilor, conflictul a escaladat dintr-o negociere de termeni într-un episod de „disciplinare instituțională”. Trei pârghii apar recurent în descrieri [9, 10, 20, 11]:

- Excluderea prin etichetare – clasificarea unui furnizor ca fiind drept „riscant” nu este doar o opinie, ea poate declanșa proceduri de evitare, înlocuire, audit suplimentar și chiar interdicții în anumite rețele [6].
- Presiunea contractuală care apare atunci când guvernul își arogă dreptul de a suspenda, rezilia sau redirecționa contracte... iar în domenii strategice contractele nu sunt doar venituri, ci și validare și acces la infrastructuri.
- Pârghii legale și de mobilizare industrială folosite în situații considerate critice; iar aici, statul american are un istoric valoros al utilizării instrumentelor de mobilizare a producției și a capacităților industriale atât în domeniul civil (proiectele *Apollo*) cât și militar (proiectul *Manhattan*).

Efectul sistemic al acestor instrumente este ușor de subestimat; un furnizor de AI nu vinde doar „un model”, el vinde o promisiune de continuitate, suport, *patch*-uri de securitate, compatibilitate cu medii clasificate și, foarte important, predictibilitate [48, 49]. Dacă un furnizor este etichetat drept riscant, atunci contractorii devin reticenți în a-l folosi: nu pentru că modelul nu ar funcționa, ci pentru că riscul birocratic devine prea mare.

Aici apare o ironie: într-o logică pur tehnică, securitatea lanțului de aprovizionare ar trebui să vizeze riscuri precum: compromiterea codului, dependențe nesigure, infrastructură vulnerabilă, acces neautorizat. În conflictul de față, aceeași etichetă pare să fie folosită și ca instrument pentru a gestiona un risc de natură „politico-normativă”: furnizorul ar putea refuza, în viitor, o utilizare considerată critică.

Cu alte cuvinte, *supply-chain risk* devine o categorie care amestecă securitatea tehnică cu obedienta contractuală. Această amestecare, dacă se confirmă ca practică, este un precedent major.

Pivotul *Anthropic*: schimbarea RSP și dilema credibilității

În centrul suspiciunilor s-a aflat și un eveniment care, în ochii multora, arată rău chiar dacă ar fi fost o coincidență: publicarea unei versiuni actualizate a politicii de scalare responsabilă (RSP 3.0) exact în proximitatea escaladării disputei. Presa internațională a remarcat că noua versiune a politicii ar fi formulat altfel promisiunile: mai puțin ca interdicții categorice, mai mult ca o combinație de evaluări, praguri și condiții [9, 30, 33].

Aici totuși merită să fim exacți în analiză. O politică de siguranță poate evolua legitim din două motive:

- Maturizare: pe măsură ce sistemele devin mai complexe, „promisiunile absolute" devin greu de menținut fără a cădea în ipocrizie. Companiile pot încerca să înlocuiască sloganuri cu proceduri.
- Presiune competitivă: într-o cursă tehnologică, un laborator poate considera că oprirea unilaterală nu reduce riscul global, ci doar transferă avantajul către competitori mai puțin scrupuloși.

În același timp, există o problemă de credibilitate: dacă identitatea publică a unei companii a fost construită pe ideea „nu scalăm fără siguranță", iar apoi textul devine mai flexibil tocmai când crește presiunea politică, publicul poate interpreta schimbarea ca o capitulare. Chiar dacă intern schimbarea ar fi justificată tehnic, extern ea arată ca o adaptare oportunistă.

Un sceptic ar formula întrebarea astfel: dacă „firul roșu" este real, de ce trebuie rescris exact când este testat? Iar un alt sceptic, situat de data asta pe partea cealaltă, ar întreba: dacă politica rămâne rigidă în fața unui rival geopolitic care nu o respectă, nu devine ea o formă de auto-sabotaj strategic?

Adevărul este că ambele întrebări sunt la fel de incomode și, probabil, ambele au dreptatea lor.

Problema tehnică pe care politica o evită: LLM-urile, eroarea rară și mediile de mare miză

În orice discuție despre AI folosită în sistemele de apărare apare un paradox. Pe de o parte, utilitatea este evidentă: rezumare de rapoarte, triere, suport pentru analiză, interogare rapidă a bazelor de date, generare de ipoteze, traducere, sprijin logistic. Pe de altă parte, exact caracteristica care face LLM-urile spectaculoase – capacitatea de a produce text fluent, coerent, convingător – este și ceea ce le face periculoase în contexte critice: pot produce erori într-o formă convingătoare.

Din perspectiva științifică, un LLM este un sistem probabilistic care maximizează probabilitatea generării următorului *token* condiționat de context [50]. Nu este un sistem logic, nu are „adevăr" în sens epistemic, ci are regularități statistice. De aici apar două riscuri principale:

- Halucinațiile: producerea de informații false sau neverificate într-o formă persuasivă.
- Deplasarea responsabilității: oamenii tind să acorde credibilitate și autoritate [51] unui *output* fluent, mai ales sub presiune [19].

În context militar, eroarea nu este un disconfort, este un eveniment potențial catastrofal. Aici intră în scenă noțiunea de *tail risk*: chiar dacă sistemul are performanță medie excelentă,

contează ce se întâmplă în cazurile rare, neașteptate, când contextul este atipic, datele sunt incomplete, iar adversarul încearcă să inducă în eroare[2] [41].

Presa internațională a reflectat în repetate rânduri opinii ale unor foști oficiali sau cercetători implicați în programe de AI guvernamentale, care atrag atenția că LLM-urile sunt încă prea fragile pentru unele aplicații de securitate națională [52, 53]. Uneori aceste opinii sunt exprimate într-un limbaj colocvial („se pot comporta derutant"), dar în spate este o idee riguroasă: sistemele generative nu sunt, prin *design*, *fail-safe*.

În tradiția instituțiilor serioase, instrumentele folosite la decizii letale au criterii foarte stricte: predictibilitate, verificabilitate, trasabilitate, comportament controlabil [53]. Un LLM, în forma lui standard, este mai degrabă un sistem de asistență cognitivă decât un instrument de comandă. Această distincție, asistență versus comandă, ar fi trebuit să fie axul dezbaterii... în schimb, scandalul a împins discuția în zona politică: cine e „atent", cine e „radical", cine „sabotază" armata. E un simptom al epocii: dezbaterile tehnice sunt absorbite de identități.

Human in the loop **nu este doar un slogan, ci o arhitectură morală și juridică**

În tradiția dreptului conflictelor armate, există principii care au rolul de a păstra războiul, oricât de brutal, într-o anumită limită de normativitate: distincția între combatanți și necombatanți, proporționalitate, precauție. Aceste principii sunt aplicate de oameni, nu de mașini, pentru că presupun judecată contextuală, responsabilitate și posibilitatea de refuz.

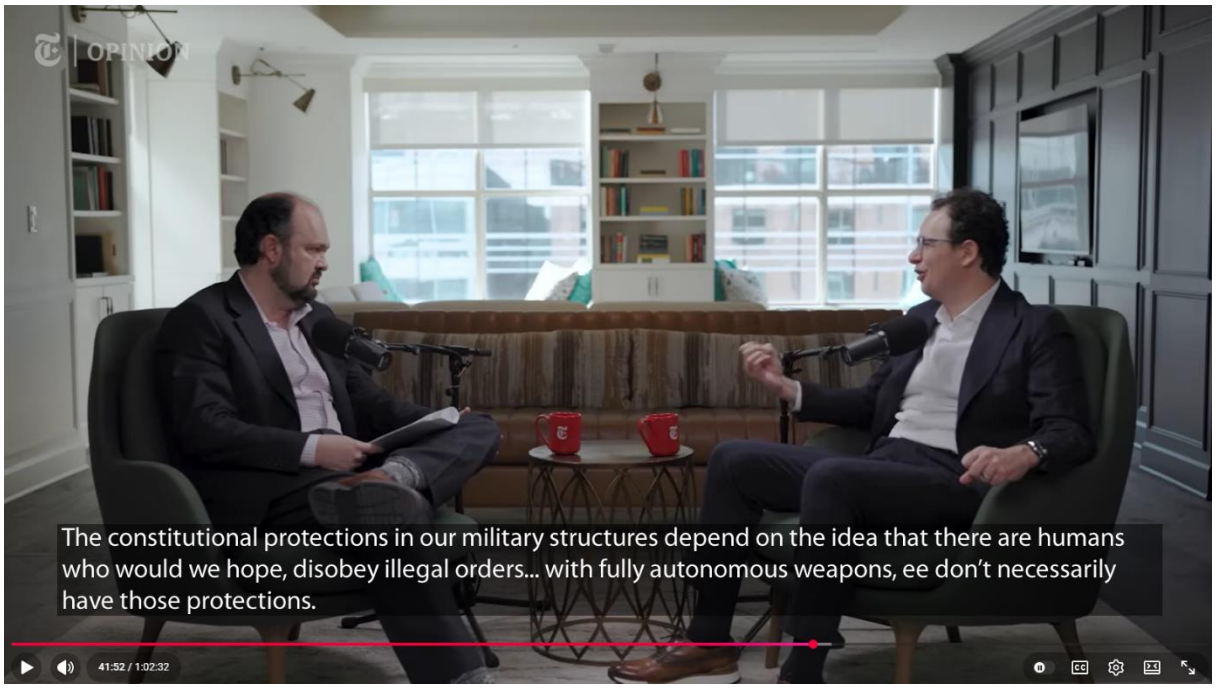
Un argument frecvent invocat în relatările presei internaționale, asociat poziției *Anthropic*, este că automatizarea completă a deciziei letale ar elimina o supapă esențială: capacitatea unui om de a opri un ordin ilegal sau disproporționat [54, 19, 55]. În sistemele militare tradiționale, ideea că „am primit ordin", nu este o scuză absolută... există răspundere [56, 41].

Când introduci un sistem autonom care selectează și angajează ținte, întrebarea devine: cine a decis? Algoritmul? Cine a semnat? Cine a înțeles? Cine a verificat? În termeni tehnici, apare problema *accountability gap* – o breșă a responsabilității.

Un critic ar putea spune: „dar armata folosește automatizare de decenii". Da, așa este, dar automatizarea tradițională, chiar în sisteme complexe, are reguli deterministe sau criterii bine definite, testate în condiții stricte, iar decizia finală, în multe arhitecturi, rămâne în mâna operatorului. Introducerea unor sisteme autonome sau a unor agenți cu autonomie largă complică dramatic trasabilitatea.

De aceea, *human in the loop* nu e un moft. Este o încercare de a păstra o structură morală și juridică veche, tocmai pentru a preveni degradarea responsabilității într-un sistem tehnologic nou.

[2] Să ne amintim când (în 1962, în timpul crizei rachetelor din Cuba) sistemele de la bordul submarinului sovietic B-59, au detectat în mod eronat un atac nuclear și, intrând în alertă, erau cât pe ce să declaseze un război nuclear. Numai prezența oamenilor (și nici aceștia prea mulți pentru că doar unul din trei comandanți, și anume Vasili Arkhipov, a refuzat să lanseze în execuție procedura de lansare a unei torpile nucleare) a salvat omenirea de la o catastrofă [65].



CEO-ul *Anthropic,* Dario Amodei (dreapta), într-un interviu cu Ross Douthat de la *New York Times.*
Sursa: https://www.youtube.com/watch?v=N5JDzS9MQYI

De ce a reacționat dur Pentagonul – suveranitate, continuitate și frica de veto privat

Dacă privim conflictul cu ochi reci, reacția dură a Pentagonului, așa cum este descrisă de presa internațională, nu este surprinzătoare. Există trei motive structurale [10, 9, 20]:

1. Suveranitate operațională. Armata nu acceptă ușor ideea că un furnizor privat poate condiționa utilizarea unui instrument în funcție de valori interne. În logica militară, asta ar însemna ca lanțul comenzii să includă implicit un consiliu corporativ.
2. Continuitate și predictibilitate. În război, nu poți fi dependent de un furnizor care, la un moment politic sensibil, decide că nu mai livrează sau blochează un anumit tip de utilizare. Chiar dacă intenția furnizorului este morală, efectul poate fi strategic.
3. Standardizarea ecosistemului. Pentagonul preferă standarde și compatibilitate. Dacă fiecare furnizor vine cu propriile interdicții, se creează un peisaj fragmentat, greu de integrat și de auditat.

În această lumină, etichetarea de tip *supply-chain risk* poate fi înțeleasă ca o formă de a spune: „nu putem construi infrastructură critică pe un furnizor care își rezervă dreptul de veto”. Este un argument pragmatic, nu neapărat moral [21, 41, 10].

Dar pragmatismul militar are limite; în societățile democratice, tocmai tradiția controlului civil și a legalității ar trebui să împiedice pragmatismul să se transforme în arbitrar. De aceea, întrebarea esențială revine: dacă statul cere libertate totală, cine garantează că această libertate nu va produce abuzuri sau accidente catastrofale?

De la negociere la „război cultural” sau cum ajunge să se piardă subiectul real

Un episod semnificativ, relatat de presa internațională, este intrarea politicului în scenă cu o retorică polarizantă: acuzații de ideologie, etichete și apel la indignare publică. Când discuția despre arme autonome și supraveghere domestică devine o discuție despre „atenție” versus „patriotism”, se întâmplă două lucruri [41]:

- Se simplifică o problemă complexă până la caricatură.
- Se substituie analiza tehnică și juridică cu identități (și abordări) tribale.

Aceasta este o degradare a dezbaterii publice, nu o victorie pentru vreuna dintre părți. Dacă *Anthropic* are dreptate în privința riscurilor autonomiei letale, retorica politică îi poate discredita argumentul. Dacă Pentagonul are dreptate în privința suveranității și continuității, retorica poate masca întrebările legitime despre control și abuz.

În mod paradoxal, scandalul ar putea avea un efect invers celui dorit de actorii politici: să întărească tendința companiilor de a se ascunde în limbaj tehnic și de a evita explicitarea valorilor, de teamă să nu fie ținte. Ori, într-un domeniu cu risc major, evitarea explicitării valorilor este periculoasă: lipsa de claritate nu înseamnă neutralitate; înseamnă că valorile vor fi impuse de cel mai puternic.

Industria și precedentul *Maven* – lecția că *Silicon Valley* nu este un contractor clasic

Presa internațională a făcut frecvent o paralelă cu episodul *Project Maven*, când mii de angajați *Google* au protestat împotriva colaborării cu Pentagonul pentru analiză video în teatre de operațiuni [57, 58]. Lecția acelui episod pentru *Washington* a fost dureroasă: companiile din *Silicon Valley* au o cultură internă care poate bloca proiecte militare, iar reputația publică contează [59, 23, 23, 60].

În 2026, diferența este că AI generativă nu e doar un instrument de analiză ci poate deveni infrastructură cognitivă: sistem de sinteză, de planificare, de recomandare, eventual de „agent" care execută pași [30]. De aceea, conflictul cu *Anthropic* a fost interpretat ca un test pentru întreaga industrie: pot laboratoarele de cercetare AI să mențină limite normative când statul tratează tehnologia ca activ strategic [16]?

În același timp, industria este competitivă. Dacă un furnizor este împins afară, altul îi poate lua locul. Presa internațională a făcut relatări inclusiv despre mișcări rapide ale altor actori majori pentru a umple vidul. Aceasta pune presiune pe orice companie care ar vrea să rămână „purist": dacă refuzi, competitorul acceptă, iar tehnologia ajunge oricum în ecosistemul militar – doar că printr-un furnizor poate mai puțin preocupat de siguranță [10, 21].

Aici apare dilema morală a laboratoarelor de cercetare de top:

- dacă accepți fără limite, riști să devii complice la abuz;
- dacă refuzi complet, riști să pierzi influența asupra modului în care AI va fi folosită, lăsând spațiul altora.

Este o dilemă reală, nu o dramă de PR.

Scandalul *Anthropic* vs. Pentagon ca simptom al unei epoci aflate între frică și oportunitate

La suprafață, scandalul *Anthropic*–Pentagon este o ceartă despre termeni contractuali și despre cine decide. În profunzime, este o confruntare între două paradigme [19, 9, 10]:

- paradigma statului securitar, care cere instrumente, viteză și libertate de utilizare în limitele legii;

- paradigma laboratorului de cercetare de top, care spune că riscurile pot depăși cadrul legal existent și cere fir roșu pentru a preveni abuzuri sau accidente ireversibile.

Cine are dreptate? Dacă întrebarea e pusă astfel, răspunsul devine aproape imposibil, pentru că fiecare parte are o dreptatea ei. Statul are dreptate când cere suveranitate și continuitate. Compania are dreptate când spune că legalitatea nu este suficientă într-o tehnologie care poate amplifica puterea de decizie și supraveghere.

De aceea, adevărata întrebare nu este „cine câștigă". Este: ce fel de ordine normativă construim pentru AI militară și guvernamentală? Dacă ordinea va fi stabilită prin presiune și etichete, vom avea o tehnologie puternică și un control slab. Dacă ordinea va fi stabilită doar prin termeni privați, vom avea veto corporativ și fragmentare.

Tehnologia e nouă. Dar principiile care împiedică degradarea puterii sunt vechi, încercate și, tocmai de aceea, indispensabile.

References

[1] The New Yorker, „What Is Claude? Anthropic Doesn't Know, Either," 16 02 2026. [Interactiv]. Available: [1] "What Is Claude? Anthropic Doesn't Know, Either," The New Yorker, Feb. 16, 2026. [Online]. Available: https://www.newyorker.com/magazine/2026/02/16/what-is-claude-anthropic-doesnt-know-either. [Accesat 01 03 2026].

[2] TechCrunch, „Claude: Everything you need to know about Anthropic's AI," 25 02 2025. [Interactiv]. Available: https://techcrunch.com/2025/02/25/claude-everything-you-need-to-know-about-anthropics-ai/. [Accesat 01 03 2026].

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

[4] C. Vrabie, „Programarea 2.0 – Cum transformă AI procesul de creare a aplicațiilor," *All in on Tech (AIoT),* vol. 1, nr. 1, 2025.

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

[6] Reuters, „Anthropic says it will challenge Pentagon's supply chain risk designation in court," 28 02 2026. [Interactiv]. Available: https://www.reuters.com/world/us/anthropic-says-it-will-challenge-pentagons-supply-chain-risk-designation-court-2026-02-28/. [Accesat 01 03 2026].

[7] Associated Press, „What to know about the clash between the Pentagon and Anthropic over military's AI use," 28 02 2026. [Interactiv]. Available: https://apnews.com/article/0c464a054359b9fdc80cf18b0d4f690c. [Accesat 01 03 2026].

[8] The Boston Globe, „Pentagon assault on Anthropic sends shock waves across Silicon Valley," 28 02 2026. [Interactiv]. Available: https://www.bostonglobe.com/2026/02/28/nation/pentagon-assault-on-anthropic-sends-shockwaves-across-silicon-valley/. [Accesat 01 03 2026].

[9] Anthropic, „Anthropic's Responsible Scaling Policy: Version 3.0," 24 02 2026. [Interactiv]. Available: https://www.anthropic.com/news/responsible-scaling-policy-v3. [Accesat 01 03 2026].

[10] Fortune, „OpenAI sweeps in to ink deal with Pentagon as Anthropic is designated a 'supply chain risk'—an unprecedented action likely to crimp its growth," 28 02 2026. [Interactiv]. Available: https://fortune.com/2026/02/28/openai-pentagon-deal-anthropic-designated-supply-chain-risk-unprecedented-action-damage-its-growth/. [Accesat 01 03 2026].

[11] Federal News Network, „Trump orders US agencies to stop using Anthropic technology in clash over AI safety," 02 27 2026. [Interactiv]. Available: https://federalnewsnetwork.com/artificial-intelligence/2026/02/anthropic-refuses-to-bend-to-pentagon-on-ai-safeguards-as-dispute-nears-deadline/. [Accesat 01 03 2026].

[12] Wired, „Anthropic Hits Back After US Military Labels It a 'Supply Chain Risk," 27 02 2026. [Interactiv]. Available: https://www.wired.com/story/anthropic-supply-chain-risk-shockwaves-silicon-valley/. [Accesat 01 03 2026].

[13] The Guardian, „OpenAI to work with Pentagon after Anthropic dropped by Trump over company's ethics concerns," 28 02 2026. [Interactiv]. Available: https://www.theguardian.com/technology/2026/feb/28/openai-us-military-anthropic. [Accesat 01 03 2026].

[14] M. Atif și A. Alamgir, „A Comprehensive Framework for Enhancing National Security through AI," *Journal of Political Stability Archive,* vol. 3, nr. 3, pp. 1443-1462, 2025.

[15] M. S. Bexheti, „The Role of Artificial Intelligence in Foreign Policy and International Security," *International Journal of Law and Politics Studies,* vol. 7, nr. 6, pp. 01-05, 2025.

[16] CNBC, „Trump admin blacklists Anthropic as AI firm refuses Pentagon demands,” 27 02 2026. [Interactiv]. Available: https://www.cnbc.com/2026/02/27/trump-anthropic-ai-pentagon.html. [Accesat 01 03 2026].

[17] The Verge, „Anthropic refuses Pentagon’s new terms, standing firm on lethal autonomous weapons and mass surveillance,” 27 02 2026. [Interactiv]. Available: https://www.theverge.com/news/885773/anthropic-department-of-defense-dod-pentagon-refusal-terms-hegseth-dario-amodei. [Accesat 01 03 2026].

[18] The Verge, „Defense secretary Pete Hegseth designates Anthropic a supply chain risk,” 28 02 2026. [Interactiv]. Available: https://www.theverge.com/policy/886632/pentagon-designates-anthropic-supply-chain-risk-ai-standoff. [Accesat 01 03 2026].

[19] The New York TImes, „‘The Business of War’: Google Employees Protest Work for the Pentagon,” 08 04 2018. [Interactiv]. Available: https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html. [Accesat 01 03 2026].

[20] The Hill, „Anthropic calls supply chain risk designation ‘unprecedented,’ ‘legally unsound’,” 27 02 2026. [Interactiv]. Available: https://thehill.com/policy/technology/5759929-pentagon-anthropic-supply-chain-risk/. [Accesat 01 03 2026].

[21] TechCrunch, „Pentagon moves to designate Anthropic as a supply-chain risk,” 27 02 2026. [Interactiv]. Available: https://techcrunch.com/2026/02/27/pentagon-moves-to-designate-anthropic-as-a-supply-chain-risk/. [Accesat 01 03 2026].

[22] C. Vrabie, „Convergenta securitatii digitale,” *Smart Cities International Conference (SCIC) Proceeding,* vol. 3, pp. 267-277, 2015.

[23] Wired, „3 Years After the Project Maven Uproar, Google Cozies to the Pentagon,” 10 11 2021. [Interactiv]. Available: https://www.wired.com/story/3-years-maven-uproar-google-warms-pentagon/. [Accesat 01 03 2026].

[24] The Washington Post, „You are hardwired to blindly trust AI. Here's how to fight it,” 03 05 2025. [Interactiv]. Available: https://www.washingtonpost.com/technology/2025/06/03/dont-trust-ai-automation-bias/. [Accesat 01 03 2026].

[25] Business Insider, „Anthropic drops signature safety pledge after conflict with Pentagon,” 25 02 2026. [Interactiv]. Available: https://www.businessinsider.com/anthropic-changing-safety-policy-2026-2. [Accesat 01 03 2026].

[26] The New Yorker, „Can Anthropic Control What It's Building?,” 11 02 2026. [Interactiv]. Available: https://www.newyorker.com/podcast/political-scene/anthropic-is-in-a-race-ai-against-its-own-ai. [Accesat 01 03 2026].

[27] The Sun, „What is Claude AI and who funds it?,” 11 09 2025. [Interactiv]. Available: https://www.thesun.co.uk/tech/36617608/what-is-claude-ai-app-anthropic/. [Accesat 01 03 2026].

[28] Contrary Research, „Anthropic,” 06 02 2026. [Interactiv]. Available: https://research.contrary.com/company/anthropic. [Accesat 01 03 2026].

[29] Kitrum, „The Inspiring Journey of Daniela Amodei, Anthropic’s Leader,” 03 07 2024. [Interactiv]. Available: https://kitrum.com/blog/the-inspiring-story-of-daniela-amodei-anthropics-leader/. [Accesat 01 03 2026].

[30] ETIH, „Anthropic updates Responsible Scaling Policy as AI risk debate shifts,” 26 02 2026. [Interactiv]. Available: https://www.edtechinnovationhub.com/news/anthropic-updates-responsible-scaling-policy-as-ai-risk-debate-shifts. [Accesat 01 03 2026].

[31] Sky News, „Google staff protest against company's work with Pentagon drone programme,” 05 04 2018. [Interactiv]. Available: https://news.sky.com/story/goggle-staff-protest-against-companys-work-with-pentagon-drone-programme-11317327. [Accesat 01 03 2026].

[32] Jacobin, „Tech Workers Versus the Pentagon,” 06 2018. [Interactiv]. Available: https://jacobin.com/2018/06/google-project-maven-military-tech-workers. [Accesat 01 03 2026].

[33] CBS News, „Hegseth declares Anthropic a supply chain risk, restricting military contractors from doing business with AI giant,” 28 02 2026. [Interactiv]. Available: https://www.cbsnews.com/news/hegseth-declares-anthropic-supply-chain-risk/. [Accesat 01 03 2026].

[34] M. G. Jacobides, S. Brusoni și F. Candelon, „The Evolutionary Dynamics of the Artificial Intelligence Ecosystem,” *Strategy Science,* vol. 6, nr. 4, 2021.

[35] D. G. Widder, M. Whittaker și S. M. West, „Why ‘open’ AI systems are actually closed, and why this matters,” *Nature,* vol. 635, pp. 827-833, 2024.

[36] M. Steidl, M. Felderer și R. Ramler, „The pipeline for the continuous development of artificial intelligence models—Current state of research and practice,” *Journal of Systems and Software,* vol. 199, 2023.

[37] N. Kolli, J. W. Sajja și A. Nerella, „Building Secure AI Agents for Autonomous Data Access in Compliance/Regulatory-Critical Environments,” *Computer Fraud and Security,* nr. 9, 2024.

[38] J. M. Schraagen, „Responsible use of AI in military systems: prospects and challenges,” *Ergonomics,* vol. 66, pp. 1719-1729, 2023.

[39] T. Pace și B. Ranes, „Bias, explainability, transparency, and trust for AI-enabled military systems,” în *Proceedings Volume 13054, Assurance and Security for AI-enabled Systems*, Maryland, 2024.

[40] N. Generous, B. Cook și J. Pruet, „Preserving security in a world with powerful AI Considerations for the future Defense Architecture," *ArXiv,* 2025.

[41] Politico, „Trump orders all federal agencies to cease using Anthropic," 27 02 2026. [Interactiv]. Available: https://www.politico.com/news/2026/02/27/trump-orders-all-federal-agencies-to-stop-using-anthropic-00804517. [Accesat 01 03 2026].

[42] S. Haack, „Monopoly on the Use of Force," în *The Bonn Handbook of Globality*, Springer, 2019, pp. 1107-1115.

[43] M. Jachtenfuchs, „The monopoly of legitimate force: denationalization, or business as usual," *European Review,* vol. 13, nr. S1, 2005.

[44] J. Schuett, „Risk Management in the Artificial Intelligence Act," *European Journal of Risk Regulation,* vol. 15, pp. 367-385, 2024.

[45] B. Judge, M. Nitzberg și S. Russell, „When code isn't law: rethinking regulation for artificial intelligence," *Policy and Society,* vol. 44, nr. 1, pp. 85-97, 2025.

[46] W. L. Currie, J. M. Leimeister și L. Willcocks, „Rethinking technology regulation in the age of AI risks," *Journal of Information Technology,* vol. 40, nr. 3, 2025.

[47] E. Zaidan și I. A. Ibrahim, „AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective," *Humanities and social sciences communications,* vol. 11, 2024.

[48] M. Haakman, L. Cruz, H. Huijgens și A. v. Deursen, „AI lifecycle models need to be revised," *Empirical Software Engineering,* vol. 26, 2021.

[49] W. Hummer, V. Muthusamy, T. Rausch, P. Dube, K. E. Maghraoui și A. Murthi, „ModelOps: Cloud-Based Lifecycle Management for Reliable and Trusted AI," în *2019 IEEE International Conference on Cloud Engineering (IC2E)*, Prague, 2019.

[50] 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], Bucharest: Pro Universitaria, 2024.

[51] C. Vrabie, „Fenomenul AI psychosis și posibilele efecte ale interacțiunii cu ChatGPT," *All in on Tech (AIoT),* vol. 1, nr. 1, 2025.

[52] W. N. Caballero și P. R. Jenkins, „On Large Language Models in National Security Applications," *Stat,* vol. SI, 2025.

[53] The Alan Turing Institute, „The Rapid Rise of Generative AI. Assessing risks to safety and security," Centre for Emerging Technology and Security, 2023.

[54] Council of Foreign Relations, „Anthropic and Pentagon Clash," 27 02 2026. [Interactiv]. Available: https://www.cfr.org/articles/anthropic-and-pentagon-clash. [Accesat 01 03 2026].

[55] Interesting Times with Ross Douthat, „Anthropic's CEO: 'We Don't Know if the Models Are Conscious'," 12 02 2026. [Interactiv]. Available: https://www.youtube.com/watch?v=N5JDzS9MQYI. [Accesat 01 03 2026].

[56] npr, „OpenAI announces Pentagon deal after Trump bans Anthropic," 28 02 2026. [Interactiv]. Available: https://www.npr.org/2026/02/27/nx-s1-5729118/trump-anthropic-pentagon-openai-ai-weapons-ban. [Accesat 01 03 2026].

[57] The New York TImes, „Pentagon Gives A.I. Company an Ultimatum," 24 02 2026. [Interactiv]. Available: https://www.nytimes.com/2026/02/24/us/politics/pentagon-anthropic.html. [Accesat 01 03 2026].

[58] Wired, „Google Won't Renew Controversial Pentagon AI Project," 02 06 2018. [Interactiv]. Available: https://www.wired.com/story/google-wont-renew-controversial-pentagon-ai-project/. [Accesat 01 03 2026].

[59] LessWrong, „Responsible Scaling Policy v3," 24 02 2026. [Interactiv]. Available: https://www.lesswrong.com/posts/HzKuzrKfaDJvQqmjh/responsible-scaling-policy-v3. [Accesat 01 03 2024].

[60] Axios, „Employees urge Google to drop Pentagon AI project," 04 04 2018. [Interactiv]. Available: https://www.axios.com/2018/04/04/employees-urge-google-to-drop-p-1522861843. [Accesat 01 03 2026].

[61] Axios, „Pentagon's use of Claude during Maduro raid sparks Anthropic feud," 13 02 2026. [Interactiv]. Available: https://www.axios.com/2026/02/13/anthropic-claude-maduro-raid-pentagon. [Accesat 01 03 2026].

[62] The Wall Street Journal, „See How the U.S. Military Captured Maduro in a Matter of Hours," 04 01 2026. [Interactiv]. Available: https://www.wsj.com/world/americas/see-how-the-u-s-military-captured-maduro-in-a-matter-of-hours-ee87179c. [Accesat 01 03 2026].

[63] The Wall Street Journal, „Pentagon Used Anthropic's Claude in Maduro Venezuela Raid," 15 02 2026. [Interactiv]. Available: https://www.wsj.com/politics/national-security/pentagon-used-anthropics-claude-in-maduro-venezuela-raid-583aff17. [Accesat 01 03 2026].

[64] Kelo by Thomson Reuters, „US used Anthropic's Claude during the Venezuela raid, WSJ reports," 13 02 2026. [Interactiv]. Available: https://kelo.com/2026/02/13/us-used-anthropics-claude-during-the-venezuela-raid-wsj-reports/. [Accesat 01 03 2026].

[65] Agerpres, „DOCUMENTAR: 20 ani de la dezastrul submarinului nuclear Kursk (12 august),” 04 08 2020. [Interactiv]. Available: https://agerpres.ro/documentare/2020/08/04/documentar-20-ani-de-la-dezastrul-submarinului-nuclear-kursk-12-august--554238. [Accesat 01 03 2026].