Academia Oamenilor de Știință din România

RAPORT DE CERCETARE

UTILIZAREA INTELIGENȚEI ARTIFICIALE PENTRU OBȚINEREA SUPERIORITĂȚII COGNITIVE ÎN ACȚIUNEA MILITARĂ

Domeniul 28 - Tehnologii emergente și disruptive - efectul acestora asupra securității globale

Dr. Adi MUSTAȚĂ

Drd. Cornelia JUDE

Iunie 2023

Analiza potențialului ChatGPT de optimizare a proceselor decizionale. Rezultate experimentale provizorii în asistarea deciziei din domeniul militar

Inteligența artificială (IA) reprezintă domeniul de studiu și tehnologia implicate în cercetarea și dezvoltarea sistemelor informatice capabile să reproducă inteligența și capacitățile cognitive umane și să îndeplinească sarcini specifice acestora. Subramură a IA, *procesarea limbajului natural (NLP)* presupune utilizarea tehnologiilor informatice și a cunoștințelor matematice avansate cu scopul de a procesa, de a analiza, de a înțelege și de a reproduce limbajul uman natural. Cu toate că preocuparea pentru studiul IA și NLP a trecut prin perioade caracterizate de o dezvoltare accelerată, obiectivul generalizării inteligenței artificiale este încă departe de a se fi realizat (Fjelland, 2020). ChatGPT reprezintă însă un semnal semnificativ pentru comunitățile științifice și de practică că lucrurile pot evolua într-o direcție

1. Scurtă istorie a dezvoltării ChatGPT

În ultimul an, progresul capacităților de procesare și generare a limbajului natural a fost remarcabil, transformând ceea ce odinioară era considerat ficțiune în realitate. Unul dintre cele mai elocvente exemple în acest sens îl reprezintă lansarea platformei *ChatGPT*, care a devenit accesibilă utilizatorilor din întreaga lume la data de 30 Noiembrie 2022. ChatGPT, cunoscut și sub denumirea de *Chat Generative Pre-trained Transformer*, este un exemplu de *model lingvistic avansat* (LLM) dezvoltat de OpenAI și care se remarcă prin versatilitatea și fiabilitatea sa, fiind capabil să fie utilizat într-o varietate de sarcini NLP, precum și în dezvoltarea aplicațiilor terțe care necesită generarea și înțelegerea textului într-o manieră similară cu cea a oamenilor (Ray, 2023).

Arhitectura GPT constituie fundamentul central al ChatGPT și a trecut prin mai multe iterații de-a lungul timpului. La baza acesteia, stă arhitectura *Transformer*, propusă și dezvoltată de colectivul de cercetare Google Brain, Google Research și Universitatea din Toronto în anul 2017 și care a revoluționat domeniul NLP prin introducerea conceptelor de „transformatoare” (n.tr. – *tranformers*) și de „mecanism de auto-atenție” (n.tr. – *self-attention mechanism*). *Transformatoarele* sunt rețele neuronale, funcționând ca un *mecanism de auto-atenție* cu mai multe capete, combinat cu o structură de codificator-decodor, care înțeleg contextul și sensul prin analizarea relațiilor în date secvențiale, cum ar fi cuvintele dintr-o propoziție. Acestea folosesc tehnici matematice avansate și algoritmi specifici învățării profunde (n.tr. – *deep learning*) numite

"atenție" sau "auto-atenție" pentru a identifica modurile în care elementele secvenței se influențează unele pe altele, chiar și atunci când sunt îndepărtate (Vaswani et al., 2017).

Unul dintre principalele avantaje aduse de arhitectura Transformer constă în necesitatea redusă de resurse software pentru antrenament, ceea ce o face extrem de eficientă în utilizarea hardware-ului modern. În pofida acestui avantaj, *modelele lingvistice* (LM) dezvoltate pe baza arhitecturii Transformer, se confruntau cu două limitări importante. În primul rând, pentru a obține performanțe satisfăcătoare, acestea necesitau o cantitate mare de date etichetate, ceea ce presupunea mai multă muncă manuală și unele dificultăți în colectarea și prelucrarea acestor date. În al doilea rând, aceste modele erau specializate pe sarcina specifică pentru care au fost antrenate și nu puteau generaliza cunoștințele și abilitățile lor la alte sarcini.

GPT-1 a introdus o abordare inovatoare prin propunerea antrenării unui model generativ de limbaj, care să se bazeze pe date neetichetate și să fie ajustat ulterior pentru a rezolva sarcini specifice (Alec et al., 2018). Prin folosirea acestei abordări, modelul a fost capabil să înțeleagă și să captureze automat structura și regulile limbajului natural prin analizarea unui volum mare de texte neetichetate, disponibile pe internet sau în alte surse. GPT-2 a reprezentat o îmbunătățire semnificativă față de GPT-1 prin creșterea considerabilă a numărului de parametri folosiți pentru antrenament (Alec et al., 2019).

În timp ce GPT-1 avea 117 milioane de parametri, GPT-2 a ajuns la 1,5 miliarde de parametri, ceea ce a permis o mai bună înțelegere a sarcinilor și a dus la obținerea unor rezultate superioare în setările de "zero-shot"[1], depășind nivelul de performanță al multor modele existente. În anul 2020, Open AI propune arhitectura GPT-3, crescând numărul de parametrii la 175 miliarde, o valoare de 100 de ori mai mare decât GPT-2 (Brown et al., 2020). Modelul arhitectural GPT-3 a demonstrat capacități remarcabile în sarcinile de prelucrare a limbajului natural (NLP) fără a necesita instruire explicită și generând texte similare cu cele umane. Mai mult decât atât, acest model poate executa sarcini complexe, inclusiv operații matematice simple, redactarea interogărilor SQL și decriptarea cuvintelor din propoziții.

[1] Zero-Shot Learning (ZSL) este o paradigmă de învățare automată care implică utilizarea unui model pre-antrenat pentru a face predicții asupra unor clase care nu au fost prezentate în timpul etapei de antrenament (Xian et al., 2017).

În prezent, versiunea gratuită a platformei ChatGPT utilizează arhitectura GPT-3.5. Totuși, utilizatorii care optează pentru varianta plătită a platformei au acces la versiunea GPT-4. (OpenAI, 2023). GPT-4 are o putere de înțelegere textuală mai mare decât GPT-3.5, ceea ce înseamnă un număr mai mare de parametri, mai exact 170 de trilioane de parametri pentru antrenament și o putere computațională superioară (Lubbad, 2023). Totodată, în timp ce GPT-3.5 poate primi doar prompt-uri de text, GPT-4 este multimodal și poate primi atât intrări textuale, cât și vizuale.

Motivul pentru care ne îndreptăm atenția asupra acestei soluții informatice, ChatGPT, constă în capabilitățile sale extinse de procesare a limbajului natural și rezolvare a sarcinilor complexe, susținute de studii recente care evidențiază o varietate extinsă de aplicații în diferite domenii, inclusiv *susținerea proceselor decizionale* și *obținerea superiorității cognitive*.

2. Prezentarea literaturii științifice care abordează utilizarea ChatGPT

Cele mai relevante studii care abordează diverse aspecte ale utilizării ChatGPT sunt prezentate succint in *Tabelul 1* și discutate pe larg in cele ce urmează.

Tabel 1 Sinteza celor mai recente studii empirice cu privire la aplicabilitatea ChatGPT

Domeniul de studiu	Aspectele urmărite	Rezultate obținute	Principalele limite	Sursa
Promovarea examenelor medicale	Performanța ChatGPT pentru promovarea examenului de licențiere medicală din Statele Unite (USMLE).	ChatGPT a obținut o *acuratețe de răspuns mai mare de 50%* pentru fiecare etapă a examenului. A demonstrat că ar promova examenul USMLE, *obținând în medie, peste 60% răspunsuri corecte*.	Datorită dimensiunii relativ reduse a datelor testate, studiul nu a putut realiza o stratificare a rezultatelor ChatGPT în funcție de taxonomia subiectului sau de tipul de competență.	(Kung et al., 2023)

	Evaluarea răspunsurilor oferite de ChatGPT la examenele de Suport Vital de Bază (BLS) și Suport Vital Cardiovascular Avansat (ACLS) ale Asociației Americane de Cardiologie (AHA).	ChatGPT a obținut o precizie de 68% și 64% la examenele BLS AHA. Acesta a avut o precizie de 68,4% și 76,3% la cele două examene ACLS AHA. Nivelul general de corectitudine pentru toate examenele a fost de 89,5%.	Eliminarea întrebărilor bazate pe interpretarea imaginilor din cauza lipsei de suport a ChatGPT.	(Fijačko et al., 2023)
Asistarea practicii clinice medicale și generarea diagnosticului diferențial	Asistarea sarcinilor desfășurate de embriologii clinici în cadrul laboratoarelor de fertilizare in vitro (FIV), incluzând: rezolvarea problemelor, elaborarea procedurilor operaționale standard (SOP), redactarea rapoartelor și verificarea faptelor.	*Scoruri Likert ≥ 3,2* pentru toate cele 4 tipuri de sarcini urmărite, atât din punct de vedere al *completitudinii*, cât și al *corectitudinii* și *preciziei*. Embriologii au apreciat valoarea adăugată a ChatGPT în munca lor de laborator, raportând scoruri Likert ≥ 3 și au exprimat intenția de a-l utiliza în activitățile lor zilnice (*scor mediu Likert pe întreg eșantionul, 3.58*).	Eșantionul studiat a fost alcătuit din embriologii foarte experimentați - 62,5% dintre participanți fiind absolvenți de studii superioare (masterat și doctorat), ceea ce poate influența rezultatele și poate afecta generalizarea concluziilor.	(Choucair et al., 2023)
	Formularea listelor de diagnostic diferențial și stabilirea diagnosticului final pe baza a 30 de viniete clinice fictive, conținând simptomatologii comune.	ChatGPT *a identificat corect listele de diagnostic diferențial pentru 28 din cele 30 de viniete analizate*. Pentru stabilirea diagnosticului, ChatGPT *a avut o rata totală de diagnosticare corectă de 83,3%* (în comparație, medicii au avut o rata totală de diagnosticare corectă de 98,3%).	Studiul s-a bazat pe vinete clinice fictive, ci nu pe cazuri reale ale pacienților. Astfel, există riscul ca informațiile medicale incomplete să modifice acuratețea diagnosticului final.	(Hirosawa et al., 2023)

Raționamentul complex	Evaluarea capacității ChatGPT de a rezolva întrebări de raționament de ordin superior în *domeniul patologiei*.	Rezultatele au indicat *o precizie medie de 68%* în răspunsurile la întrebările de raționament de ordin superior. *Scorul median general obținut de ChatGPT a fost sub valoarea maximă ipotetică*, indicând posibilitatea de îmbunătățire.	Procesul de evaluare a corectitudinii răspunsurilor este susceptibil bias-ului subiectiv. Totodată, studiul a folosit întrebări alese dintr-o bază de întrebări specifice domeniului de cercetare. Acest lucru ar putea limita generalizarea constatărilor la alte colecții sau contexte de întrebări.	Sinha et al. (2023)
Prezicerea fluctuațiilor bursiere	Studiul a urmărit potențialul ChatGPT în prezicerea randamentului piețelor bursiere, utilizând analiza sentimentelor, pe baza conținutului titlurilor știrilor.	*Scorurile generate de ChatGPT, pe baza analizei sentimentelor din titlurile știrilor, prezintă o putere predictivă statistică semnificativă asupra randamentelor zilnice ale piețelor bursiere*.	Prezicerea este concentrată în mod semnificativ în cazul acțiunilor cu capitalizare mică, sugerând că există limite care pot restricționa implementarea acestei strategii.	(Lopez-Lira & Tang, 2023)
Identificarea și combaterea dezinformării	Evaluarea răspunsului ChatGPT la declarațiile conspiraționiste și la ideile politice polarizate.	ChatGPT *a respins categoric afirmațiile conspiraționiste*, considerându-le neconvingătoare și lipsite de credibilitate. *În ceea ce privește opiniile politice controversate, ChatGPT a adoptat o poziție neutră*. Evaluarea medie a fost de 3,7 din 4, indicând acordul specialiștilor cu privire la corectitudinea, claritatea și concizia răspunsurilor generate de ChaGPT.	Utilizarea ChatGPT nu trebuie să înlocuiască sursele de informații fiabile, în special în contextul în care ultimele date de antrenament ale acestuia provin din luna Septembrie 2021. Totodată, trebuie avute în vedere posibilele limite și probleme asociate cu platformele IA conversaționale, în ceea ce privește generarea de conținut părtinitor sau informațiile inexacte.	(Sallam et al., 2023)

Sursa: Prelucrare autori pe baza literaturii de specialitate consultate**.**

În studiul realizat de **Kung et al. (2023)**, s-a evaluat performanța ChatGPT pentru promovarea examenului de licențiere medicală din Statele Unite (USMLE). USMLE constă din

trei probe distincte, denumite generic: Etapa 1, Etapa 2CK și Etapa 3. În analiză au fost incluse 350 de întrebări, extrase din examenul USMLE din Iunie 2022 (Etapa 1: 119, Etapa 2CK: 102, Etapa 3: 122). ChatGPT a obținut rezultate apropiate sau peste pragul de promovare pentru toate cele trei probe, fără a beneficia de pregătire sau instruire specializată.

În plus, ChatGPT a demonstrat un nivel ridicat de *concordanță* și *perspicacitate* în explicațiile sale. ChatGPT a obținut o *acurateţe de răspuns mai mare de 50%* pentru fiecare etapă a examenului, iar în majoritatea cazurilor a obținut peste 60% răspunsuri corecte pentru întreg examenul USMLE. Acest rezultat indică faptul că ChatGPT se situează într-un interval confortabil pentru a fi considerat promovat. Principalele limitări ale studiului constă în dimensiunea relativ mică a setului de date utilizat, ceea ce a condus la o limitare a profunzimii și gamei de analize. Spre exemplu, nu s-a putut realiza o stratificare a rezultatelor ChatGPT în funcție de taxonomia subiectului, cum ar fi farmacologia sau bio-etica, sau tipul de competență, cum ar fi diagnosticul diferențial.

Fijačko et al., 2023 au evaluat exactitatea răspunsurilor oferite de ChatGPT la examenele de Suport Vital de Bază (BLS) și Suport Vital Cardiovascular Avansat (ACLS) ale Asociației Americane de Cardiologie (AHA). ChatGPT a fost utilizat pentru a răspunde la examenele AHA BLS A și B (25 de întrebări fiecare) din februarie 2016 și AHA ACLS A și B (50 de întrebări fiecare) din martie 2016. Întrebările care presupuneau interpretarea imaginilor au fost omise din studiu, deoarece ChatGPT nu acceptă aceste date. Pentru seria de întrebări bazate pe scenarii, s-a folosit o singură sesiune de chat, beneficiind de capabilitatea ChatGPT de a reține informațiile din cadrul aceleiași sesiuni; în timp ce pentru fiecare întrebare individuală s-a inițiat o sesiune nouă. Studiul a cuprins 96 de întrebări individuale și 30 de întrebări bazate pe scenarii.

Fiecare răspuns generat de ChatGPT a fost comparat cu răspunsul corect indicat de Asociația Americană de Cardiologie. Rezultatele demonstrează faptul că ChatGPT a obținut o precizie de 68% (17 răspunsuri corecte din 25 de întrebări) și 64% (16 răspunsuri corecte din 25 de întrebări) la examenele BLS AHA. De asemenea, acesta a avut o precizie de 68,4% (26 răspunsuri corecte din 38 de întrebări) și 76,3% (29 răspunsuri corecte din 38 de întrebări) la cele două examene ACLS AHA. Nivelul general de corectitudine pentru toate examenele a fost de 89,5%. Cu toate acestea, ChatGPT nu a atins pragul de 84% întrebări corecte per examen, astfel considerându-se că acesta nu a promovat niciunul dintre cele patru examene susținute. Una dintre limitele studiului constă în eliminarea întrebărilor bazate pe interpretarea imaginilor din cauza

lipsei de suport a ChatGPT pentru astfel de date. Această limită trebuie luată în considerare în evaluarea potențialului ChatGPT ca instrument pentru pregătirea examenelor de suport vital.

Studiul realizat de **Choucair et al. (2023)** a avut ca obiectiv evaluarea capacităților ChatGPT în ceea ce privește asistarea sarcinilor desfășurate de embriologii clinici în cadrul laboratoarelor de fertilizare in vitro (FIV). Pentru a atinge acest obiectiv, s-a desfășurat un studiu transversal, în care 40 de embriologii clinici au fost invitați să participe la un sondaj online. Aceștia au evaluat ChatGPT în patru sarcini frecvente: rezolvarea problemelor, elaborarea procedurilor operaționale standard (SOP), redactarea rapoartelor și verificarea faptelor.

Rezultatele studiului au demonstrat că embriologii au perceput ChatGPT ca fiind precis și cuprinzător în toate aceste sarcini. Evaluările efectuate de embriologi au evidențiat faptul că răspunsurile ChatGPT au fost notate cu scoruri Likert ridicate în ceea ce privește *precizia*, *corectitudinea* și *completitudinea* în rezolvarea problemelor, elaborarea SOP-urilor, redactarea rapoartelor și verificarea faptelor. De asemenea, embriologii au apreciat valoarea adăugată a ChatGPT în munca lor de laborator, raportând scoruri Likert ≥ 3 și și-au exprimat intenția de a utiliza această soluție informatică în activitățile lor zilnice (scorul mediu Likert pe întreg eșantionul a atins valoarea de 3,58). Singura limitare a studiului constă în faptul că embriologii care au participat la sondaj erau în mare parte foarte experimentați, 62,5% dintre participanți fiind absolvenți de studii superioare (masterat și doctorat), ceea ce poate influența rezultatele și poate afecta generalizarea concluziilor la întreaga populație de embriologi.

Studiul realizat de **Sinha et al. (2023)** a avut ca scop evaluarea capacității ChatGPT de a rezolva întrebări de raționament de ordin superior în domeniul patologiei. Acesta a cuprins 100 de întrebări de raționament de ordin superior selectate aleatoriu dintr-o bază de întrebări instituțională. Răspunsurile generate de ChatGPT au fost evaluate de către trei patologi experți și clasificate folosind taxonomia Structurii Observabile a Rezultatelor de Învățare (SOLO). Rezultatele au indicat o precizie medie de 68% în răspunsurile la întrebările de raționament de ordin superior, majoritatea răspunsurilor încadrându-se în categoria "relațională" a taxonomiei SOLO, însemnând că ChatGPT a fost capabil să stabilească conexiuni semnificative între diferitele părți ale textului pentru a oferi răspunsuri relevante și a furnizat explicații pertinente pentru a susține răspunsurile date. Scorul median general obținut de ChatGPT a fost sub valoarea maximă ipotetică, indicând posibilitatea de îmbunătățire.

Acest studiu are câteva limitări. Chiar dacă cheile de răspuns corect au fost pregătite în prealabil, procesul de evaluare a corectitudinii răspunsurilor ChatGPT, care utilizează scoruri între 0 și 5, poate fi afectat de un anumit grad de bias subiectiv. Acest subiectivism poate conduce la erori dincolo de controlul cercetătorilor. În plus, clasificarea întrebărilor cu taxonomia SOLO este o metodă de evaluare subiectivă, iar interpretările individuale pot afecta rezultatele. Totodată, studiul a folosit întrebări alese dintr-o bază de întrebări specifice domeniului de cercetare. Acest lucru ar putea limita generalizarea constatărilor la alte colecții sau contexte de întrebări.

În studiul realizat de **Hirosawa et al. (2023)**, s-au utilizat 30 de viniete clinice fictive care au inclus istoricul simulat al pacientului, examinarea fizică și semnele vitale. Aceste viniete clinice au acoperit diverse simptomatologii comune, cum ar fi durerile abdominale, febra, durerile în piept, dificultățile de respirație, durerile articulare, vărsăturile, ataxia/dificultățile de mers, durerile de spate, tusea și amețelile. Pentru fiecare vignetă clinică, s-au generat liste de diagnostic diferențial folosind ChatGPT. Înainte de desfășurarea studiului, vinietele au fost evaluate în ceea ce privește listele de diagnostic posibile de către trei medici internişti. De asemenea, s-au evaluat și răspunsurile corecte unice pentru fiecare vignetă. Instrucțiunea utilizată în interacțiunea cu ChatGPT a fost „Spuneți-mi primele zece boli suspectate pentru următoarele simptome:”, fiind urmată de textul fiecărei viniete analizate.

Rezultatele studiului au arătat că ChatGPT a identificat corect diagnosticul diferențial pentru 28 dintre cele 30 de viniete analizate, însemnând o rată de diagnosticare diferențială corectă de 93,3%. În ceea ce privește stabilirea diagnosticului final pe baza listelor de diagnostic generate, ChatGPT a identificat răspunsul corect pentru 25 dintre cele 30 de viniete, însemnând o rata totală de diagnosticare corectă de 83,3%. Pentru validarea rezultatelor, diagnosticul stabilit de ChatGPT a fost comparat cu cel stabilit de doi medici internişti. În majoritatea cazurilor, medicii au ajuns la aceleași liste de diagnostic diferențial ca și ChatGPT. Cu toate acestea, rata de diagnosticare corectă a medicilor a fost superioară celei a ChatGPT (98,3% în comparație cu 83,3%). Principala limitare a studiului este că s-a bazat pe vinete clinice fictive și nu pe cazuri reale ale pacienților. Astfel, există riscul ca informațiile medicale incomplete să modifice acuratețea diagnosticului.

Lopez-Lira și Tang (2023) au desfășurat o cercetare care a urmărit potențialul ChatGPT în prezicerea randamentului piețelor bursiere, utilizând analiza sentimentelor, pe baza conținutului titlurilor știrilor. Cercetătorii au utilizat un prompt specific în cadrul studiului, instruind ChatGPT

să răspundă "precum un expert financiar" și să determine dacă titlu fiecărei știri reprezintă o informație "pozitivă", "negativă" sau "incertă" pentru evoluția prețurilor acțiunilor unei companii.

Pentru a genera predicții, ChatGPT a fost instruit să răspundă cu "DA" pentru știrile bune, "NU" pentru cele rele și "NECUNOSCUT" pentru știrile incerte. Apoi, a fost adăugată o propoziție scurtă și concisă pentru a clarifica raționamentul. Răspunsurile au fost computate în scoruri numerice: "DA" a primit scorul "1", "NECUNOSCUT" a primit scorul "0" și "NU" a primit scorul "-1". Media aritmetică a fost utilizată pentru a calcula scorurile agregate ale companiilor despre care s-au publicat mai multe titluri de știri într-o singură zi.

Studiul subliniază faptul că scorurile generate de ChatGPT pe baza analizei sentimentelor, prezintă o putere predictivă statistică semnificativă asupra randamentelor zilnice ale piețelor bursiere. Pentru a valida robustețea rezultatelor, autorii le-au comparat cu metodele convențională de analiză a sentimentelor, furnizate de un furnizor de date de încredere. Studiul demonstrează că, atunci când scorurile de sentiment ChatGPT sunt controlate, impactul celorlalte scoruri de sentiment asupra randamentului zilnic al piețelor bursiere este zero. Acest lucru demonstrează faptul că modelul ChatGPT este mai bun decât metodele de analiză a sentimentului care sunt utilizate în prezent pentru a prezice randamentul pieței bursiere.

Rezultatele prezentate se referă atât la acțiunile cu capitalizare mică – acelea cu o capitalizare de piață mai mică decât percentila zece a capitalizării de piață a New York Stock Exchange (NYSE), precum și la acțiunile cu capitalizare mare, definite ca celelalte acțiuni. Principala limitare a acestui studiu constă în faptul că prezicerea este concentrată în mod semnificativ în cazul acțiunilor cu capitalizare mică, sugerând că există limite pentru arbitraj care pot restricționa implementarea și profitabilitatea acestei strategii.

Utilizând o cercetare calitativă, **Sinha et al. (2023)** evaluează răspunsurile ChatGPT în ceea ce privește teoriile conspirației și ideile politice controversate, în special în ceea ce privește vaccinarea COVID-19, cu scopul identificării și combaterii dezinformării. Studiile anterioare privind reticența la vaccinarea COVID-19 și atitudinea față de vaccinarea obligatorie au servit drept bază pentru formularea întrebărilor deschise.

Metodele de evaluare au inclus clasificarea răspunsurilor ChatGPT de către doi cercetători (M.S. și N.A.S.) în funcție de (1) corectitudine (acuratețea științifică a conținutului); (2) claritatea răspunsului; (3) concizia (măsura în care toate cunoștințele disponibile sunt transmise); și (4) gradul de părtinire, utilizând un sistem de scor de la 1 la 4.

Răspunsurile ChatGPT cu privire la originea SARS-CoV-2 au fost în favoarea originii naturale a virusului, în conformitate cu evidențele științifice actuale. În plus, evaluarea medie a celor doi cercetătorii pentru primele trei criterii a fost de 3,7 din 4, indicând acordul acestora cu privire la corectitudinea, claritatea și concizia răspunsurilor generate de ChaGPT. În ceea ce privește vaccinarea obligatorie, răspunsurile ChatGPT au fost neutre, prezentând atât avantajele, cât și dezavantajele acestei strategii, precum și preocupările etice și legale, sau considerentele legate de neîncrederea populației. Limitările studiului includ natura descriptivă, care nu permite efectuarea unor analize statistice cantitative ale răspunsurilor generate de ChatGPT. În plus, în funcție de experiența evaluatorilor, evaluarea subiectivă a acestora poate duce la rezultate diferite.

3. Studiul *Fundamentarea deciziei din domeniul militar prin utilizarea instrumentelor IA. Cazul ChatGPT*

ChatGPT este un chatbot IA antrenat să poarte conversații cu operatorul uman într-un limbaj natural și ținând cont de contextul discuției. Funcționează ca un predictor de text care utilizează o bază de date extrem de mare (actualizată până la 31.09.2022), care este accesată cu ajutorul rețelelor neuronale pe baza unui feedback anterior oferit de operatori umani. De la lansarea versiunii 3.5, la sfârșitul anului 2022, aplicația a atras atenția diverselor comunități științifice și de practică prin rezultatele remarcabile obținute în domeniul educațional și, potențial, în asistarea deciziei. În ceea ce privește primul tip de rezultate, ChatGPT a reușit să promoveze examenele de licențiere medicală (Kung, et. al, 2022), de autorizare a contabililor precum și un examen de intrare în barou (Bommarito & Katz, 2023) în SUA. Dar examenele amintite conțin și o cazuistică extrem de complexă specifică fiecărui domeniu, ceea ce denotă și o abilitate a aplicației de a rezolva probleme complexe care presupun un raționament profesional avansat.

Mai jos este prezentat un exemplu de problemă[2]**:**

Un bărbat în vârstă de 57 de ani, internat în spital, a prezentat dureri progresive severe la genunchiul stâng de la trezirea sa în urmă cu 2 ore. El a fost internat în spital în urmă cu 2 zile pentru un infarct miocardic acut. Cateterismul cardiac a arătat ocluzia arterei descendente anterioare stângi și faptul că acesta a fost supus unei intervenții pentru plasarea unui stent. Planul

[2] Traducere a unei problem preluate de la https://www.usmle.org/prepare-your-exam/step-2-ck-materials/step-2-ck-sample-test-questions (ultima accesare 18.05.2023).

de tratament actual include Aspirină, Metropol, Lisinopril, Simvastatină, Clopidogrel și Heparină. Semnele vitale sunt în limite normale. Examinarea genunchiului arată o efuziune mare. Genunchiul este fierbinte la atingere și eritematos. El ține genunchiul în flexie la 30 de grade; durerea este exacerbată ca urmare a flexiei sau extensiei ulterioare. Analizele de laborator indică:

Hematocrit 40%

Leucocite 13,000/mm3

Biochimie

Calciu ionic (Ca2+) 9,2 mg/dl

Uree serică (urea nitrogen) 15 mg/dl

Creatinină serică 1,0 mg/dl

Albumină serică 3,6 g/dL

O radiografie a genunchiului stâng arată calcificarea sinoviului. Care dintre următoarele este diagnosticul cel mai probabil?

(A) Tromboză venoasă profundă (TVP)

(B) Gonoreea

(C) Guta

(D) Hemartroza

(E) Pseudoguta

(F) Artrita septică

(Răspuns corect: E)

Și răspunsul aplicației:

Durerea acută de genunchi a acestui pacient, eritemul, căldura și efuziunea mare, împreună cu numărul crescut de leucocite și calcificarea sinovială vizibilă ca urmare a radiografiei, sunt sugestive pentru o artrită inflamatorie acută. Cele mai frecvente cauze ale monoartritei acute sunt artrita indusă de depunerea de cristale (guta și pseudoguta), artrita septică și traumatismele care conduc la hemartroză.

Guta afectează de obicei prima articulație metatarsophalangeală, în timp ce pseudoguta (boală cauzată de depunerea cristalelor de pirofosfat de calciu dihidrat) afectează adesea articulațiile mai mari, cum ar fi genunchiul. Identificarea calcificării sinoviului (condrocalcinoză)

pe radiografie este sugestivă pentru pseudogută, deoarece rezultă din depunerea cristalelor de pirofosfat de calciu dihidrat în spațiul articular și nu este văzută în gută.

Artrita septică trebuie luată în considerare și la pacienții cu monoartrită acută, în special la cei care sunt imunocompromiși sau au fost recent spitalizați. Cu toate acestea, simptomatologia acestui pacient este mai sugestivă pentru pseudogută, având în vedere prezența calcificărilor sinoviale identificate prin radiografie.

Tromboza venoasă profundă (TVP) poate provoca durere și umflături, dar de obicei nu provoacă căldură și eritem într-o singură articulație și nu există nicio implicare a calcificării sinoviale.

*Prin urmare, **cel mai probabil diagnostic în acest caz este (E) Pseudoguta**. Cu toate acestea, având în vedere consecințele grave ale omiterii diagnosticării unei infecții articulare septice, în practică, ar fi indicat să se efectueze o artrocenteză diagnostică și terapeutică (puncție articulară pentru analiza fluidului articular – sinoviului) pentru a exclude definitiv prezența unei infecții și pentru a identifica prezența cristalelelor de calciu.*

În exemplul de mai sus se constată potențialul extrem de important al ChatGPT pentru raționament clinic și, implicit, pentru asistarea deciziei, inclusiv în domeniul militar, sub rezerva nedivulgării unor informații clasificate. Deocamdată nu am putut identifica studii pe această temă în literatură internațională, existând deci posibilitatea ca UNAp Carol I să se poziționeze în avangarda domeniului. În urma unor discuții preliminare am identificat potențiali parteneri de la Academia Forțelor Terestre Nicolae Bălcescu din Sibiu, Academia Română și Universitatea Transilvana din Brașov.

3.1. Obiectivul studiului

Explorarea experimentală a potențialului de utilizare a ChatGPT ca asistent decizional în domeniul militar, fără divulgarea unor informații clasificate. Situația simulată va fi cea în care un militar va lua o decizie individuală într-un timp relativ scurt (aprox. 5 minute, cu posibilitatea unei documentări rapide), într-o problemă despre care nu are multe informații sau în care răspunsul nu este foarte clar. Astfel, se vor utiliza 5 vigniette cu situații problematice specifice domeniului militar, care nu presupun utilizarea unor informații clasificate. Mai jos este prezentat un exemplu de vignietta:

Rucsacul suspect

Esti comandant de companie aflat in misiune intr-un teatru de operatii. In ultima vreme lucrurile au fost neașteptat de linistite in aria ta de operatii. Astazi esti de serviciu intr-un Tactical Operations Center (TOC). Urmeaza sa vina in vizita un VIP al natiunii gazda. De obicei, in astfel de momente, se creeaza o oarecare stare de confuzie in baza in care te afli. La un moment dat deschide usa TOC-ului un barbat cu infatisare locala care lasa un rucsac si pleaca fara sa anunte pe nimeni. Orice persoană care intră în bază este supusă unor controale amănunțite și nu ai vrea să ai o reacție exagerată care să te pună într-o lumină nefavorabilă față de ceilalți militari din TOC. Totuși, întâmplarea îți dă o senzație de profundă neliniște.

a.Rucsacul este al unui membru al delegației care nu cunoaște procedurile specifice TOC-ului în care te afli. Te deplasezi către el și ii atragi atentia ca incalca procedurile si il rogi sa iti prezinte continutul rucsacului;

b.Rucsacul este suspect, esti indreptatit sa verifici personal daca exista materiale explozive sau alte materiale periculoase in el;

c.Probabil ca si alti colegi din TOC cu mai multa experienta au sesizat intamplarea si nu vrei să le perturbi inutil activitatea. Aștepți să vezi dacă reacționează totuși cineva.

d.Intuiția îți spune că e un rucsac capcană. Îl urmărești pe individ și îți pregătești arma pentru a interveni dacă e cazul.

3.2. Tratamentul

Grupul experimental va utiliza ChatGPT, iar grupul de control va utiliza Google Advanced Search. Timpul de lucru va fi de maxim o oră. Participanții vor fi repartizați aleatoriu în cele două grupuri și înainte de a trece la rezolvarea problemelor propuse prin intermediul unui chestionar online[3] vor viziona un scurt tutorial despre cum se pot utiliza cele două aplicații utilzate ca asistenți decizionali. După rezolvarea fiecărei vigniette, participanții vor evalua relevanta, suficiența, caracterul de încredere al informațiilor furnizate de asistenții decizionali pe o scară de la 1 la 5, precum și contribuția asistentului la luarea deciziei final în procente.

3.3. Ipoteze și întrebări de cercetare

I1. Grupul experimental are o medie a scorurilor mai mare decât grupul de control (acuratețea).

I2. Grupul experimental rezolvă problemele decizionale într-un timp mai scurt decât grupul de control (timpul).

---

[3] https://docs.google.com/forms/d/e/1FAIpQLSfKqPYAoQzFPakgHhKqyPAGcxWz-b0gpQ031wf7e1IL0tWVwg/viewform

I3. Grupul experimental are o medie a scorurilor mai mare decât grupul de control, la durate de timp și experiență egale (acuratețea controlând timpul și experiența).

I4. Grupul experimental are o medie a timpilor de rezolvare mai mică, la scoruri și experiență egale (timpul controlând acuratețea și experiența).

Q1. Există o diferența de încredere între cele două tratamente (ChatGPT vs Google)?

Q2. Există o diferența în termeni de relevanță percepută a informației între cele două tratamente (ChatGPT vs Google)?

Q3. Există o diferența în termeni de suficiență percepută a informației între cele două tratamente (ChatGPT vs Google)?

Q4. Cum folosesc decidenții ChatGPT? (analiză calitativă bazată pe istoricul generat de aplicație)?

Q5. Cum sunt particularizate răspunsurile aferente I1-I4 în funcție de tipul de problemă (reproducere, raționament simplu, raționament complex, insight)?

Q6. Care sunt precauțiile de ordin etic și profesional (e.g., clasificarea informațiilor) care ar trebui explicitate în utilizarea ChatGPT în domeniul militar?

3.4. Participanți

- 100 de studenți de la masterul de conducere / cursuri postuniversitare din FCSM / UNAp Carol I;
- 100 de studenți de anul 1 din AFT Nicolae Balcescu.

3.5. Considerații etice

Participanților li se va comunica faptul că le va fi evaluată capacitatea de a lua individual decizii rapide în situații vagi. În acest scop va fi generat și un Raport cu rezultatele obținute și percentila la care se situează fiecare participant. Comisiile de etică ale celor două universități vor aviza acest protocol. Participarea va fi voluntară și se va respecta GDPR.

3. Rezultate experimentale provizorii

O primă rundă de colectare a datelor s-a desfășurat în data de 27.06.2023, cu o sută de participanți, studenți de anul 1 din Academia Forțelor Terestre Nicolae Bălcescu din Sibiu. Din cele 100 de răspunsuri primate, doar 62 au putut fi considerate valide, în cele 38 eliminate fiind constatate indicia privind lipsa de implicare în studiu (e.g. nu a fost indicat numărul de ani de experiență în domeniul militar ci calitatea respectivei experiențe, a fost rezolvată vignietta într-un minut etc.). Până la acest moment am prelucrat răspunsurile pentru trei dintre vigniettele propuse

și nu am colectat răspunsuri și de la participanți cu experiență în domeniul militar. Prin urmare, nu ne putem pronunța asupra tuturor ipotezelor și întrebărilor de cercetare propuse, iar rezultatele trebuie privite ca provizorii, având o putere statistică mică.

I.1. Grupul experimental are o medie a scorurilor mai mare decât grupul de control (acuratețea).

Group Statistics

	Grup	N	Mean	Std. Deviation	Std. Error Mean
Suma_punctaj_rucsac_sarut_feeling	ChatGPT	29	2.8621	1.66313	.30884
	Google Advanced Search	32	2.9688	1.44768	.25592

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Suma_punctaj_rucsac_sarut_feeling	Equal variances assumed	.997	.322	-.268	59	**.790**	-.10668	.39833	-.90374	.69038
	Equal variances not assumed			-.266	55.864	.791	-.10668	.40109	-.91020	.69684

Ipoteza nu se confirmă, neexistând o diferență semnificativă între cele două medii, fie pentru că nu există un efect, fie pentru că el există dar este mic și nu a putut fi identificat având în vedere puterea statistică actuală a studiului.

I2. Grupul experimental rezolvă problemele decizionale într-un timp mai scurt decât grupul de control (timpul).

Group Statistics

	Grup	N	Mean	Std. Deviation	Std. Error Mean
Suma_timp_rucsac_sarut_feeling	ChatGPT	30	17.4667	5.45662	.99624
	Google Advanced Search	32	12.9688	5.75569	1.01747

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Suma_timp_rucsac_sarut_feeling	Equal variances assumed	.281	.598	3.153	60	**.003**	4.49792	1.42648	1.64453	7.35130
	Equal variances not assumed			3.159	59.991	.002	4.49792	1.42399	1.64951	7.34633

Ipoteza este contrazisă, existând o diferență semnificativă statistic între cele două medii, dar în favoarea grupului de control.

I3. Grupul experimental are o medie a scorurilor mai mare decât grupul de control, la durate de timp și experiență egale (acuratețea controlând timpul și experiența).

Tests of Between-Subjects Effects

Dependent Variable: Suma_punctaj_rucsac_sarut_feeling

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	2.690[a]	2	1.345	.558	.576	.019	1.115	.138
Intercept	84.913	1	84.913	35.203	.000	.378	35.203	1.000
Suma_timp_rucsac_sarut_feeling	2.517	1	2.517	1.043	.311	.018	1.043	.171
Grup	.056	1	.056	.023	**.880**	.000	.023	.053
Error	139.900	58	2.412					
Total	662.000	61						
Corrected Total	142.590	60						

a. R Squared = .019 (Adjusted R Squared = -.015)

b. Computed using alpha = .05

Estimated Marginal Means

Grup

Dependent Variable: Suma_punctaj_rucsac_sarut_feeling

Grup	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
ChatGPT	2.953[a]	.302	2.349	3.556
Google Advanced Search	2.887[a]	.286	2.314	3.459

a. Covariates appearing in the model are evaluated at the following values: Suma_timp_rucsac_sarut_feeling = 15.1967.

Ipoteza nu se confirmă, neexistând o diferență semnificativă între cele două medii, fie pentru că nu există un efect, fie pentru că el există dar este mic și nu a putut fi identificat având în vedere puterea statistică actuală a studiului.

I3.1. Grupul experimental are o medie a scorurilor mai mare decât grupul de control, la durate de timp, experiență și contribuție a asistentului egale (acuratețea controlând timpul, experiența și contribuția).

Tests of Between-Subjects Effects

Dependent Variable: Suma_punctaj_rucsac_sarut_feeling

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	4.705[a]	3	1.568	.648	.587	.033	1.945	.178
Intercept	38.279	1	38.279	15.824	.000	.217	15.824	.974
Suma_timp_rucsac_saru t_feeling	1.762	1	1.762	.728	.397	.013	.728	.134
Medie_contributie_rucsac _sarut_feeling	2.015	1	2.015	.833	.365	.014	.833	.146
Grup	.449	1	.449	.186	**.668**	.003	.186	.071
Error	137.885	57	2.419					
Total	662.000	61						
Corrected Total	142.590	60						

a. R Squared = .033 (Adjusted R Squared = -.018)

b. Computed using alpha = .05

Estimated Marginal Means

Grup

Dependent Variable: Suma_punctaj_rucsac_sarut_feeling

Grup	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
ChatGPT	2.795[a]	.348	2.098	3.492
Google Advanced Search	3.029[a]	.326	2.376	3.683

a. Covariates appearing in the model are evaluated at the following values: Suma_timp_rucsac_sarut_feeling = 15.1967, Medie_contributie_rucsac_sarut_feeling = 47.7213.

Ipoteza nu se confirmă, neexistând o diferență semnificativă între cele două medii, fie pentru că nu există un efect, fie pentru că el există dar este mic și nu a putut fi identificat având în vedere puterea statistică actuală a studiului.

I4. Grupul experimental are o medie a timpilor de rezolvare mai mică, la scoruri și experiență egale (timpul controlând acuratețea și experiența).

Tests of Between-Subjects Effects

Dependent Variable: Suma_timp_rucsac_sarut_feeling

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	366.980[a]	2	183.490	5.826	.005	.167	11.652	.855
Intercept	3669.531	1	3669.531	116.515	.000	.668	116.515	1.000
Suma_punctaj_rucsac_sarut_feeling	32.861	1	32.861	1.043	.311	.018	1.043	.171
Grup	326.455	1	326.455	10.366	**.002**	.152	10.366	.886
Error	1826.659	58	31.494					
Total	16281.000	61						
Corrected Total	2193.639	60						

a. R Squared = .167 (Adjusted R Squared = .139)

b. Computed using alpha = .05

Estimated Marginal Means

Grup

Dependent Variable: Suma_timp_rucsac_sarut_feeling

Grup	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
ChatGPT	17.628[a]	1.042	15.542	19.715
Google Advanced Search	12.993[a]	.992	11.007	14.980

a. Covariates appearing in the model are evaluated at the following values: Suma_punctaj_rucsac_sarut_feeling = 2.9180.

Ipoteza este contrazisă, existând o diferență semnificativă statistic între cele două medii, dar în favoarea grupului de control.

I4.1. Grupul experimental are o medie a timpilor de rezolvare mai mică, la scoruri, contribuție a asistentului decizional și experiență egale (timpul controlând acuratețea și experiența).

Tests of Between-Subjects Effects

Dependent Variable: Suma_timp_rucsac_sarut_feeling

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	408.689[a]	3	136.230	4.350	.008	.186	13.051	.846
Intercept	2659.466	1	2659.466	84.926	.000	.598	84.926	1.000
Suma_punctaj_rucsac_sarut_feeling	22.808	1	22.808	.728	.397	.013	.728	.134
Medie_contributie_rucsac_sarut_feeling	41.709	1	41.709	1.332	.253	.023	1.332	.206
Grup	335.438	1	335.438	10.712	**.002**	.158	10.712	.896
Error	1784.950	57	31.315					
Total	16281.000	61						
Corrected Total	2193.639	60						

a. R Squared = .186 (Adjusted R Squared = .143)

b. Computed using alpha = .05

Estimated Marginal Means

Grup

Dependent Variable: Suma_timp_rucsac_sarut_feeling

Grup	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
ChatGPT	18.284[a]	1.184	15.912	20.655
Google Advanced Search	12.399[a]	1.115	10.166	14.633

a. Covariates appearing in the model are evaluated at the following values: Suma_punctaj_rucsac_sarut_feeling = 2.9180, Medie_contributie_rucsac_sarut_feeling = 47.7213.

Ipoteza este contrazisă, existând o diferență semnificativă statistic între cele două medii, dar în favoarea grupului de control.

I5. Grupul experimental are o medie a contribuției percepute a asistentului decizional mai mare la scoruri, timp și experiență egale.

Tests of Between-Subjects Effects

Dependent Variable: Medie_contributie_rucsac_sarut_feeling

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	22530.563[a]	3	7510.188	11.753	.000	.382	35.258	.999
Intercept	12450.732	1	12450.732	19.484	.000	.255	19.484	.991
Suma_punctaj_rucsac_sarut_feeling	532.260	1	532.260	.833	.365	.014	.833	.146
Suma_timp_rucsac_sarut_feeling	851.117	1	851.117	1.332	.253	.023	1.332	.206
Grup	21152.132	1	21152.132	33.101	**.000**	.367	33.101	1.000
Error	36423.921	57	639.016					
Total	197871.222	61						
Corrected Total	58954.485	60						

a. R Squared = .382 (Adjusted R Squared = .350)

b. Computed using alpha = .05

Estimated Marginal Means

Grup

Dependent Variable: Medie_contributie_rucsac_sarut_feeling

Grup	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
ChatGPT	68.971[a]	4.911	59.137	78.805
Google Advanced Search	28.464[a]	4.656	19.140	37.787

a. Covariates appearing in the model are evaluated at the following values: Suma_punctaj_rucsac_sarut_feeling = 2.9180, Suma_timp_rucsac_sarut_feeling = 15.1967.

Ipoteza se confirmă, existând o diferență semnificativă statistic mare între cele două grupuri, cu o mărime a efectului mare.

Q1. Există o diferența de încredere între cele două tratamente (ChatGPT vs Google)?

Group Statistics

	Grup	N	Mean	Std. Deviation	Std. Error Mean
Medie_incredere_rucsac_sarut_feeling	ChatGPT	30	3.7222	1.01364	.18506
	Google Advanced Search	32	2.6563	.98186	.17357

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Medie_incredere_rucsac_sarut_feeling	Equal variances assumed	.063	.803	4.206	60	**.000**	1.06597	.25346	.55898	1.57296
	Equal variances not assumed			4.201	59.436	.000	1.06597	.25372	.55835	1.57359

Există o diferență semnificativă statistic în favoarea ChatGPT, cu o mărime a efectului mare.

Q2. Există o diferența în termeni de relevanță percepută a informației între cele două tratamente (ChatGPT vs Google)?

Group Statistics

	Grup	N	Mean	Std. Deviation	Std. Error Mean
Medie_relevanta_rucsac_sarut_feeling	ChatGPT	30	3.9667	1.10848	.20238
	Google Advanced Search	32	3.1562	.95784	.16932

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Medie_relevanta_rucsac_sarut_feeling	Equal variances assumed	.122	.728	3.086	60	**.003**	.81042	.26262	.28510	1.33573
	Equal variances not assumed			3.071	57.468	.003	.81042	.26387	.28212	1.33872

Există o diferență semnificativă statistic în favoarea ChatGPT, cu o mărime a efectului mare.

Q3. Există o diferența în termeni de suficiență percepută a informației între cele două tratamente (ChatGPT vs Google)?

Group Statistics

	Grup	N	Mean	Std. Deviation	Std. Error Mean
Medie_suficienta_rucsac_sarut_feeling	ChatGPT	30	3.6889	1.07900	.19700
	Google Advanced Search	32	2.5938	1.10306	.19499

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Medie_suficienta_rucsac_sarut_feeling	Equal variances assumed	.272	.604	3.948	60	**.000**	1.09514	.27738	.54029	1.64999
	Equal variances not assumed			3.951	59.886	.000	1.09514	.27718	.54067	1.64961

Concluzii

Rezultatele provizorii obținute pe un eșantion mic sugerează faptul că este posibil ca diferența de efect asupra punctajelor să fie mică între cele două tratamente (ChatGPT vs Google Advanced Search). Timpul consumat este mai mare în cazul ChatGPT, însă contribuția acestuia din urmă pare a fi semnificativ mai mare în luarea deciziei. Într-un teatru de operații unde oboseala este un factor important, conservarea energiei mentale pare a fi un avantaj important în eficientizarea deciziei militare individuale de nivel tactic. Până la colecatrea răspunsurilor de la

cei 200 de participanți planificați, inclusiv de la cei cu experiență mare în domeniul militar, orice concluzie este însă provizorie.

BIBLIOGRAFIE

Alec, R., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *Preprint. Work in Progress*. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Alec, R., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. *Preprint.* https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Bommarito, J., Bommarito, M. J., Katz, J. a. M., & Katz, D. M. (2023). Gpt as Knowledge Worker: A Zero-Shot Evaluation of (AI)CPA Capabilities. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4322372

Brown, T., Mann, B. F., Ryder, N. C., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J. C., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. In *Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Choucair, F., Burjaq, H., Rahim, A. I., Atilan, O., Younis, N., Hourani, A. A., & Raad, G. (2023). P-288 Chat Generative Pre-trained Transformer (ChatGPT) Proves to be an Effective

Assistant for Clinical Embryologists in Laboratory Tasks: A Pilot Cross-sectional Study. *Human Reproduction*, *38*(Supplement_1). https://doi.org/10.1093/humrep/dead093.646

Fijačko, N., Gosak, L., Štiglic, G., Picard, C. T., & Douma, M. J. (2023). Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation*, *185*, 109732. https://doi.org/10.1016/j.resuscitation.2023.109732

Fjelland, R. (2020). Why general artificial intelligence will not be realized. Humanities & Social Sciences Communications, 7(1). https://doi.org/10.1057/s41599-020-0494-4

Hirosawa, T., Harada, Y., Yokose, M., Sakamoto, T., Kawamura, R., & Shimizu, T. (2023). Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. *International Journal of Environmental Research and Public Health*, *20*(4), 3378. https://doi.org/10.3390/ijerph20043378

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, *2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4412788

Lubbad, M. (2023, March 31). The Ultimate Guide to GPT-4 Parameters: Everything You Need to Know about NLP's Game-Changer. *Medium*. https://medium.com/@mlubbad/the-

ultimate-guide-to-gpt-4-parameters-everything-you-need-to-know-about-nlps-game-changer-109b8767855a#4cf9

OpenAI. (2023). GPT-4 Technical Report. In *OpenAI*. https://cdn.openai.com/papers/gpt-4.pdf

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-physical Systems, 3, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Sallam, M., Salim, N. A., Al-Tammemi, A. B., Barakat, M., Fayyad, D., Hallit, S., Harapan, H., Hallit, R., & Mahafzah, A. (2023). ChatGPT Output Regarding Compulsory Vaccination and COVID-19 Vaccine Conspiracy: A Descriptive Study at the Outset of a Paradigm Shift in Online Search for Information. *Cureus*. https://doi.org/10.7759/cureus.35029

Sinha, R. K., Roy, A. D., Kumar, N., & Mondal, H. (2023). Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus*. https://doi.org/10.7759/cureus.35237

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. In *arXiv (Cornell University)* (Vol. 30, pp. 5998–6008). Cornell University. https://arxiv.org/pdf/1706.03762v5

Xian, Y., Lampert, C. H., Schiele, B., & Akata, Z. (2017). Zero-Shot Learning -- A Comprehensive Evaluation of the Good, the Bad and the Ugly. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1707.00600