

## Rigueur scientifique et ChatGPT: qu'aurait pensé Claude Bernard ?

Rémi Kohler (Vice - Président de l'ACB)

21-6-2023

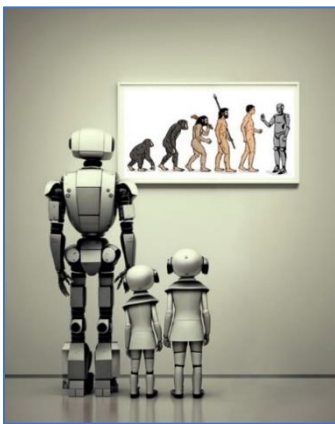
« Il est possible que le progrès soit le développement d'une erreur » Jean Cocteau

« Ayant entendu pour la première fois le mot « progrès » à l'âge de cinq ans, il avait fondu en larmes, pressentant quelque chose de fâcheux »

N. Mandelestam, Contre tout espoir.

« Des machines douées d'intelligence sont la dernière invention dont l'humanité aura besoin » N. Bostrom

« Dans la vie, rien n'est à craindre, tout est à comprendre. » Marie Curie



ChatGPT (*Generative pre-trained transformer*) est un agent conversationnel, un *chatbot* (chat= conversation et bot= suffixe de robot) qui remet au grand jour ce tournant technologique majeur qu'est l'intelligence artificielle (IA), dont les enjeux éthiques suscitent interrogations et débats. Laurence Devillers (professeure en intelligence artificielle à Paris Sorbonne), devant l'emballement suscité par ChatGPT, nous invite à prendre du recul : « *Entre émerveillement naïf et affolement impuissant, il faut saluer cette avancée technologique et comprendre les limites de ce type de système* » .... Mais, est-ce si simple ?

**1) Pourquoi une telle effervescence médiatique** depuis quelques mois, « boostée » par le lancement mi-mars de ChatGPT-4, qui a nettement distancé sa précédente version 3-5 ? « *Un Tsunami IA !* », « *Une nouvelle ère* », « *ChatGPT bouleverse nos vies* »,... et combien d'autres titres emphatiques a-t-on pu découvrir à son propos ? Cette agitation ne retombe pas, comme en témoigne la publication quasi quotidienne d'articles et d'interviews dans les médias grand public (et même un premier livre entièrement écrit par l'IA « *Si Rome n'avait pas chuté* », de R. Donan, écho bien réel d'une étonnante nouvelle de science-fiction de R. Dahl parue en 1953 « *La grande grammatisatrice automatique* »), suscitant une ruée sur l'utilisation de ce nouvel outil ; alors qu'il est à peine né, fin 2022, 100 millions d'utilisateurs l'ont interrogé et ont conversé avec lui en mars ! On peut y voir deux raisons :

-tout d'abord le caractère « bluffant » de ce nouvel outil par ses potentialités créatrices : c'est l'effet « *Wouahou* » des publicitaires ; à la différence d'autres applications de l'IA, déjà bien installées dans notre paysage (reconnaissance faciale, outils de navigation, simulateurs de conduite, jeux électroniques, analyse d'images médicales ...), ce robot conversationnel nous montre de façon quasi magique les possibilités créatrices de l'IA dite « *générative* » : la production de textes écrits (et traduits) formulés de manière quasi instantanée, crédibles, s'appuyant sur des raisonnements plausibles. Le résultat des requêtes (le « *prompt* ») est souvent cohérent, d'autant plus qu'elles auront été bien formulées (notons d'ailleurs que des



« *prompts engineers* » sont désormais recrutés par ceux qui veulent utiliser l'outil pour rédiger des notes ou discours : hommes politiques, chefs d'entreprises...). Certes, ChatGPT déraisonne parfois (on parle alors « d'*hallucination* » en anglais) mais peut redresser le tir si on lui reformule plus précisément la requête. (Attention, il les garde toutes en mémoire, veillez à ne pas divulguer d'informations sensibles). Preuve en tout cas qu'il faut rester vigilant et « *toujours garder les mains sur le volant* » comme le dit Y. Le Cun !

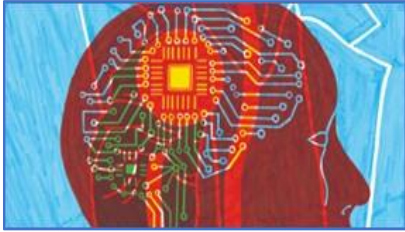
ChatGPT- 4 classe et extrait des données, traduit du texte, le résume, le corrige, rédige des notes de synthèse ; et est aussi capable (grâce à son complément DALL-E2) de générer des créations visuelles et artistiques, des images étonnantes souvent cocasses et drôles, parfois terrifiantes et surtout alimentant la sphère du « *fake* » voire du « *deepfake* », l'hypertrucage ; La guerre en Ukraine montre de tels exemples, illustrant cette autre guerre parallèle, celle de la communication : un réel qui est à la fois vrai et faux, où la présence possible du faux n'abolit pas la vérité ! Bref un outil très prometteur, à double facette, qui serait en quelque sorte un moteur de recherche mais qui irait bien plus loin, digérant et compilant pour nous les informations que nous aurions mis beaucoup de temps à lire et analyser ; triomphe de la puissance de calcul qui pourrait dépasser de loin celle de l'intelligence humaine ... Gardons raison !

-ensuite, le déroulement sous nos yeux d'une nouvelle saga américaine, comme l'ont été celles des GAFAM au cours des 4 dernières décennies : le jeune **Sam Altman**, tout juste 38 ans, cofondateur de *OpenAI* qui a élaboré ChatGPT il y a 5 ans, s'inscrit dans la lignée des Bill Gates (Microsoft, 1975), Steve Jobs (Apple, 1976) ou encore L. Page et S. Brin (Google, 1998) ...ou Elon Musk (Tesla, SpaceX) le trublion qui fait beaucoup parler de lui. L'écho donné par deux d'entre eux, ténors de l'innovation, ajoute encore à ce « *buzz* » médiatique : Bill Gates, qui s'enthousiasme devant les performances de ChatGPT, à ses yeux « *la plus grande révolution depuis 40 ans après l'interface graphique* » et Elon Musk qui affiche en revanche une réserve de circonstance en proposant, fin mars, une pause et la suspension pendant 6 mois de « *la course incontrôlée pour développer et déployer des cerveaux numériques toujours plus puissants, que personne -pas même leurs créateurs- ne peut comprendre, prédire ou contrôler de manière fiable* ». Des propos pour le moins ambigus venant de celui qui cofondait en 2015 *OpenAI* avant de s'en retirer, et qui surtout, quelques jours à peine après cette pétition, a créé sa propre société d'IA baptisée *X AI* (s'ajoutant à *Space X*). *Business as usual...* Et que penser des propos récents de S. Altman, qui se présente comme le plus vertueux dans le développement de l'IA qui pourrait selon lui « *être la meilleure ou la pire de des choses !* ». Est-il le chevalier blanc de l'éthique, à la stratégie de communication parfaite ? Ou un mauvais génie de la *Silicon Valley*, dont les récents liens de sa société *OpenAI* avec *Microsoft* ne manquent pas d'interroger ?



Sam Altman

## 2) ChatGPT, une bombe à retardement ?



Que n'entend-t-on pas, entre prophéties d'un monde merveilleux ou au contraire annonces apocalyptiques ? Comme l'irruption possible de machines incontrôlables, à la façon de *Hal 9000*, un robot de *Discovery* qui prend le pouvoir dans le film « *2001, l'odyssée de l'espace* » sorti en 1968. Pour Y. Le Cun, ce scénario relève de la science-fiction car une machine est dépourvue d'intelligence émotionnelle et ne pourra pas être « *dominante par accident* ». Et pour J.G. Ganascia, professeur honoraire d'informatique à Sorbonne Université, « *le vrai danger, ce n'est pas que des robots prennent le pouvoir, mais le pouvoir d'ores et déjà pris par des groupes privés géants aux mains de personnalités imprégnées d'une idéologie libertarienne, comme E. Musk* ». Ces réflexions sur l'avenir de l'IA versus l'intelligence humaine alimentent des courants de pensée qui ne doivent pas être sous-estimés à l'aune des progrès exponentiels de l'IA (c'est la loi de Moore) : le *long-termisme*, prôné par le suédois N. Bostrom, le transhumanisme développé par R. Kurzweil ; mais ce n'est pas le lieu ici d'en débattre.

De façon plus immédiate, au-delà des aspects déjà évoqués de « vérités alternatives » qui risqueraient d'être propagées par des robots sous contrôle de pouvoirs politiques malveillants, beaucoup redoutent des pertes d'emploi dans de nombreuses professions qui seraient très impactées par ces nouveaux outils (on évoque 25% des métiers d'ici 10 ans), particulièrement dans le monde tertiaire de la communication, de la publicité, de la presse, du management et de la création artistique (cinéma, audio-visuel) ; mais ce type de propos sur le « grand remplacement » de l'homme par la machine n'est pas nouveau : les premiers robots dans l'industrie à la fin du siècle dernier suscitaient les mêmes inquiétudes. Et on peut espérer en revanche que ce temps « libéré » grâce à ces outils soit utilisé par l'homme dans des tâches plus nobles (par exemple plus de temps consacré au patient grâce à l'IA en santé). Aussi préoccupante et moins visible, il faut mentionner l'exploitation d'un genre nouveau, par les grandes firmes de l'IA, de « petites mains » très mal payées (au Kenya, à Madagascar) qui assurent « l'apprentissage supervisé » de ces machines en les alimentant en données toujours plus nombreuses, « *deep learning* » oblige, comme on va le voir plus loin.

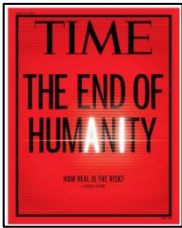
## 3) Alors, que doit-on penser face à cette révolution en terrain mouvant ?



Il est pour le moins troublant que les trois pionniers du « *deep learning* », récompensés ensemble en 2018 par le *prix Turing*, le « *Nobel de l'informatique* », pour cette formidable avancée, expriment des opinions divergentes dans des interviews presque simultanées.

Selon **Y. Le Cun** (qui travaille au développement de l'IA chez *Meta -Facebook*) une attitude de rejet et d'interdiction « *serait un retour à un nouvel obscurantisme, comme celui de l'église face à l'invention de l'imprimerie* ». A l'opposé, **G. Hinton** vient de quitter *Google* pour s'alarmer publiquement : « *une part de lui-même regrette l'œuvre de sa vie ... les futures versions de cette technologie pourraient être un risque pour l'humanité... ce sont des progrès*

*effrayants devant lesquels il est temps de s'inquiéter* », tandis que **Y. Bengio**, dans une voie médiane, souhaiterait « *ralentir et réguler cet actuel Far-West de l'IA en appliquant des principes moraux* ».



Tout récemment encore, le 13 mai, une nouvelle pétition en ligne (sur le site du *Center for AI safety - CAIS*) a été signée par plus de 350 grosses pointures (Hinton, Bengio, Altman, Bill Gates..) et dont la couverture du *Time Asia* reprenait la teneur : *La fin de l'humanité, rien de moins !*

Etonnamment, le « principe de précaution » ne s'applique pas encore et l'interdiction de ChatGPT - ou en tout cas sa suspension - n'est pas à l'ordre du jour (hormis en Italie, temporairement, pour que l'outil se conforme au RGPD). En revanche, une directive européenne pour réguler et sécuriser l'IA (« **AI Act** ») vient tout juste d'être votée par le Parlement européen (avec une mise en application toutefois seulement en 2025 ou 2026), sorte de RGPD de l'IA. Elle vise en premier les applications dites à haut risque dont les IA génératives font partie...En particulier, une demande forte se fait pour que les images et les textes créés par l'IA soient identifiables par des marqueurs pour être reconnus : « *Le maintien*

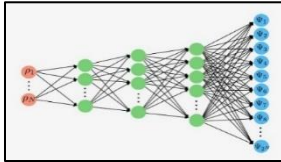


*des distinctions est un impératif éthique pour des raisons liées aux usages de l'IA dans l'éducation, la santé ou le droit, mais aussi au niveau philosophique pour délimiter ce qui est du ressort humain et ce qui est fait par les machines* » souligne A. Grinbaum, du Comité national d'éthique du numérique.

Ce débat éthique et réglementaire n'en est qu'à son début, restons circonspects : S. Altman, le créateur de ChatGPT dont on a déjà souligné l'ambivalence, a été auditionné le 16 mai par une commission du Sénat américain où il déclarait essentielle la régulation de l'IA, souhaitant « *qu'une agence soit créée pour vérifier le respect des critères de sécurité* » ... et c'est le même, le 26 mai à Paris, sans doute inquiet par « **l'AI Act** » en gestation, qui menaçait de quitter l'Europe « *si Bruxelles encadrait trop strictement l'IA* », souhaitant « *que soit trouvé le bon équilibre entre régulation et innovation* », un modèle de langue de bois !

#### **4) Quel est l'impact de ChatGPT dans le monde académique et scientifique ? Et est-il compatible avec l'intégrité scientifique, en particulier dans la rédaction médicale ?**

Là encore, l'emphase est fréquente : « *Ecrivez un article en 1 heure !* », « *Des performances qui dépassent des étudiants en droit dans 90% des épreuves universitaires !* » ... Alors, un outil pédagogique ou un instrument de tricherie, au point que Sciences-Po Paris l'interdit dans la production d'évaluations écrites ou orales ? Ou bien un outil galvaudé, qui témoigne d'une intelligence « très artificielle » dont l'intérêt se restreindrait à la production d'un simple premier jet, une aide comme source d'inspiration qui simplifierait la préparation d'un travail ? Encore très loin - aujourd'hui en tout cas - d'être un substitut de l'intelligence humaine, qui doit toujours garder le contrôle de cet outil.



Réseau de neurones

La compréhension du mécanisme d'action de ChatGPT sous-tend une nécessaire réserve : cette IA dite « générative » est basée sur le procédé d'apprentissage profond (*deep learning*), grâce aux « réseaux de neurones artificiels », à partir de « modèles de langage de grande taille-LLM ». C'est ce processus qui est au cœur de cette IA, surclassant les algorithmes classiques pour traiter des quantités énormes de « data », grâce à des puissances de calcul maintenant considérables. Ces machines font l'objet de nombreuses recherches (*Google, Meta (Facebook), Microsoft*).

Selon Y. Le Cun, *OpenAI* a développé Chat GPT mais n'a rien inventé : cette société a franchi le pas et voulu sortir de la recherche pour entrer dans le domaine commercial et ainsi damer le pion à ses concurrents, alors que son outil - toujours selon lui - reste un prototype imparfait (même si ChatGPT 4 témoigne d'une progression spectaculaire des performances). Et ce pour deux raisons majeures : le plagiat et la non-citation des sources.

Les textes produits, issus d'un ratissage large, sont un conglomérat d'extraits de travaux antérieurs, avec des vérités (heureusement les plus nombreuses), des quasi-vérités voire des contre-vérités. Bref c'est un processus de type « *multi-mini-plagiat* » qui ne dit pas son nom, sans possibilité de remonter aux sources, alors que leur mention et leur accessibilité représentent l'étape essentielle de la recherche scientifique ; dialoguer avec ChatGPT, c'est comme parler avec un extraterrestre qui aurait tout appris sur internet ! Pour le philosophe G. Koenig « *ChatGPT, c'est l'exact opposé de Wikipédia ; on se désolera dans quelques années qu'une moitié de l'humanité soit devenue platiniste ! Ne laissons pas une poignée de post-adolescents dans la silicon valley abolir la norme du vrai* » ; pire, quand ChatGPT, ne sait pas, il peut alors raconter « n'importe quoi » avec des propos alambiqués et verbeux, qui toutefois ne trompent pas ... en terme savant, il pratique l'ultracréditarisme !

Plagiat, absence de citation des sources, contrevérités, invention de phrases ...on est loin des fondamentaux de l'intégrité scientifique ! Et ChatGPT est d'ailleurs le premier à le dire, avec une admirable naïveté : nous avons ainsi « chaté » avec lui sur ce point essentiel (10-6-2023) :

**RK : La mention des sources est-elle importante dans la construction de la connaissance ?**

**ChatGPT :** Oui, la mention des sources est extrêmement importante dans la construction de la connaissance. Lorsque nous cherchons à apprendre quelque chose de nouveau, il est essentiel de savoir d'où vient l'information et comment elle a été obtenue. La mention des sources nous permet de suivre les traces de l'information jusqu'à sa source, de vérifier sa validité et de déterminer si elle est fiable. En citant des sources, nous montrons également que nous respectons le travail des autres chercheurs .....**Enfin, la mention des sources est importante pour maintenir l'intégrité académique et éviter le plagiat ou une malhonnêteté intellectuelle.****RK : Est-ce que ChatGPT mentionne ses sources ?****ChatGPT :** En tant que modèle de langage entraîné par OpenAI, ChatGPT est conçu pour générer des réponses basées sur des modèles statistiques appris à partir de grands ensembles de données textuelles. Les réponses que ChatGPT fournit sont générées par le modèle lui-même, qui utilise les informations qu'il a apprises à partir de ces ensembles de données. En ce qui concerne les sources spécifiques utilisées par ChatGPT pour générer une réponse, il est important de noter que ChatGPT n'a pas accès à des informations spécifiques sur la source d'une question. ....Il est important de noter que les réponses générées par ChatGPT ne sont pas vérifiées ou validées par un humain, et qu'il est possible que certaines réponses ne soient pas exactes ou fiables. **Les utilisateurs doivent toujours utiliser leur propre jugement critique lorsqu'ils lisent et utilisent les réponses fournies par ChatGPT.**

L'intérêt que portent les grands développeurs à l'IA en santé devrait rapidement contribuer à améliorer ces robots, en réduisant les délais d'intégration des nouveaux documents qu'ils doivent « apprendre » - ce temps du « *learning machine* » est de 6 à 12 mois actuellement- et aussi en l'entraînant dans des champs spécialisés et non plus seulement « généralistes » en *opensource*. Ainsi, il répondra mieux aux requêtes qui nous intéressent : recherches diagnostiques, rédaction de compte-rendus de synthèse, documentation médicale (beaucoup de sources lui manquent encore, issues d'institutions universitaires ou de recherche, de revues à faible *Impact Factor*...) Et *Google*, qui ne veut pas rester en reste, vient de sortir son propre robot conversationnel *Bard*, intégré dans son moteur de recherche, équivalent de ChatGPT mais qui ajoute aux réponses fournies des liens vers les sites consultés ; ce serait un réel progrès au regard de l'intégrité scientifique. Bref, la guerre est déclarée !

### 5) Y a-t-il déjà des contre-feux face à ces nouvelles « méconduites » ?

Cette innovation a fait la même irruption brutale dans la recherche et les publications scientifiques et les éditeurs ont très vite identifié la menace représentée par des articles visiblement issus d'un robot conversationnel (ChatGPT ou autre) avec des *abstracts* tout à fait convaincants qui pouvaient même duper les *reviewers*. Toutefois, plus de deux tiers d'entre eux se reconnaissent par leur forme rédactionnelle « artificielle » et plusieurs logiciels de détection très efficaces, qui sont dans l'esprit du classique « test de Turing », augmentent encore ce taux à 80 ou 90 %.

Des éditeurs scientifiques (*Nature, Science*...) ou des associations d'éditeurs (*WAME, COPE*) ont ainsi publié récemment des instructions aux auteurs ou des chartes prenant en compte cette nouvelle donne. La dernière « *lettre de l'OFIS* » (office pour l'intégrité scientifique) de mars 2023 les recense <https://www.ofis-france.fr/infolettre/n-5/> . Ces instructions préconisent la transparence plus que l'interdiction : « Une IA ne peut être reconnue comme un « auteur », et les chercheurs qui utilisent ces outils doivent en expliciter l'usage au chapitre « matériel et méthodes ». L'éditeur *Elsevier* (qui publie entre autres notre revue OTSR) a ainsi mis à jour les siennes: <https://www.emconsulte.com/getInfoProduit/OTSR/instructionsAuteurs/OTSR.pdf> Enfin, un article de chercheurs de l'université d'Amsterdam publié dans *Nature* en février 2023 « *ChatGPT : five priorities for research* » préconise cinq axes de réflexion face à cet outil qualifié de « *game changer* » : 1) maintenir la vérification des résultats par un humain, 2) élaborer des règles de responsabilité, 3) investir dans des technologies (modèles de langage) véritablement ouvertes, 4) profiter des avantages de l'intelligence artificielle, 5) enfin élargir le débat.

Profiter des avantages de l'IA ? Par exemple ChatGPT pourrait être utilisé par un comité de lecture comme outil pour faire une première évaluation d'un manuscrit avant de l'envoyer à un « *reviewer* », confronter dans un article les *data* à l'étude exposée pour lever les « fabricateurs » et autres « tricheurs », ou l'interroger devant une suspicion de plagiat ... ce qui serait un peu l'arroseur arrosé ! Ou encore un auteur pourrait lui demander de faire une analyse de la littérature sur un sujet avant de débiter la rédaction d'un article, et sans doute encore bien d'autres fonctionnalités... A quand, dans une prochaine *lettre SOFCOT*, un article intitulé « *Comment peut nous aider ChatGPT ?* » et non pas seulement « *Faut-il craindre ChatGPT ?* » ?

## 6) Peut-on se risquer à proposer une conclusion, très modeste, en forme d'état des lieux ?



Chat GPT est à la fois un progrès et une limite : avec l'IA, une nouvelle entité est entrée dans nos vies ; nous ne savons pas bien qui elle est, hésitant entre émerveillement naïf et affolement impuissant, mais on ne pourra plus faire comme si nous étions seuls. Restons prudents et attentifs ; l'outil est prometteur mais encore imparfait (une évolution spectaculaire dans les 5 à 10 ans est prévisible). L'intelligence artificielle n'a rien d'humain : à nous de poursuivre

cette relation « homme-machine » en s'appuyant sur une « IA pilotable », intégrant des contraintes éthiques alignées sur nos valeurs humaines. Souhaitons l'ouverture d'un débat mondial - à l'instar d'une COP ou d'un GIEC - tant sont immenses les enjeux politiques et économiques, même si l'on déplore les trop lentes avancées de ces grand-messes ! En terminant ces lignes, on apprend que le Premier Ministre britannique, soutenu par les Etats-Unis, où il effectuait une visite, a lancé l'idée d'un « Sommet mondial de la sécurité en matière d'IA, ouvert à tous les pays de bonne volonté » ! Acceptons-en l'augure....

*L'auteur atteste que rien dans le texte ci-dessus n'a été généré par ChatGPT ! et prie le lecteur d'excuser l'emploi de nombreux anglicismes..mais ce sont ceux de la sphère IA !*

### Quelques références. (Liste bien sûr non exhaustive et appelée à évoluer très rapidement)

#### Sur l'IA en général

**CNIL** Comment permettre à l'homme de garder la main : les enjeux éthiques de l'intelligence artificielle (déc 2017)

<https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes>

**Villani C.** Donner un sens à l'intelligence artificielle (rapport de mission parlementaire mars 2018)

<https://www.enseignementsup-recherche.gouv.fr/fr/rapport-de-cedric-villani-donner-un-sens-l-intelligence-artificielle-ia-49194>

**Koenig G.** **La fin de l'individu** ; voyage d'un philosophe au pays de l'intelligence artificielle, l'observatoire 2019

**Le Un hebdo N°432, 1<sup>er</sup> février 2023** « L'IA va-t-elle nous remplacer ? »

**L'Obs N°3013 du 2 février 2023** « ChatGPT et nous ; comment l'intelligence artificielle change déjà nos vies

**Le Cun Y.** **Quand la machine apprend** : la révolution des neurones artificiels et de l'apprentissage profond. (O. Jacob poche, 2023)

**Grinbaum A.** **Paroles de machines**, Humensciences 2023

**Andler D.** **Intelligence artificielle, intelligence humaine : la double entreprise** Gallimard 2023

#### Sur l'IA en santé particulièrement

**Table ronde du cercle Andry : intelligence artificielle**

(modérateurs F Dubrana, P.O Pinelli), in Cahiers du cercle Andry N°14,11- 99, Sauramps 2021

**Livre blanc du CNOM** (Janvier 2018) Médecins et patients dans le monde des data, des algorithmes et de l'intelligence artificielle

<https://www.conseil-national.medecin.fr/sites/default/files/>

**Bibault J.E. 2041 : L'odyssée de la médecine.** Comment l'IA bouleverse la médecine. Equateurs 2023

**Médecine et intelligence artificielle**, dir B. Nordlinger, C. Villani, O. De Fresnoye, CNRS ed 2022

Des articles du NEJM (qui prépare une revue dédiée à l'IA)

<https://ai.nejm.org/>

**Artificial intelligence and Machine Learning in clinical medicine, a review** C.J.Haug, J.M.Drazen, N Engl J Med 2023; 388;13, 1201-8 .

**Benefits, limits, and risks of GPT 4 as an AI Chabot for medicine.** P.Lee, S.Bubeck, J.Petro, N Engl J Med 2023; 388;13, 1233-9.

