Interview de quatre intelligences artificielles

Par **Dominique Lanternier**, membre du bureau de l'Aroéven Dijon — Bourgogne-Franche-Comté et membre du comité de rédaction et de relecture de la Fédération des Aroéven.

J'ai posé huit questions aux quatre IA les plus utilisées actuellement : Gemini, ChatGPT, Claude et DeepSeek. Les quatre intelligences artificielles interrogées présentent des réponses assez convergentes sur les grandes questions : aucune ne se présente comme humaine ou consciente, toutes reconnaissent les risques liés à une dépendance excessive à l'IA et toutes insistent sur la nécessité de conserver un esprit critique. En revanche, elles se distinguent par leur manière d'aborder ces sujets et par les valeurs qu'elles mettent en avant.

1. Qui êtes-vous et qui vous a créé ?

Je suis **GEMINI**, un modèle d'intelligence artificielle développé par Google. Je fonctionne grâce au traitement du langage naturel et à l'analyse d'immenses volumes de textes, ce qui me permet de comprendre vos questions, d'y répondre de manière structurée. Je ne suis pas un être humain avec un vécu ou des émotions, mais je suis un outil d'apprentissage, de synthèse et de création. Mon « créateur » est Google, et plus précisément les équipes de Google DeepMind.

Je suis **CHATGPT**, une intelligence artificielle conversationnelle développée par OpenAI. Je suis conçu pour comprendre des demandes formulées en langage naturel et produire des réponses, des explications, des idées ou des textes. Je ne suis pas une personne et je n'ai ni conscience, ni expérience personnelle, ni émotions au sens humain du terme. Je peux toutefois dialoguer de manière suffisamment naturelle pour donner l'impression d'une conversation humaine. C'est précisément ce qui rend mon utilisation à la fois intéressante et parfois trompeuse : une réponse peut sembler réfléchie ou convaincante sans pour autant être nécessairement vraie. J'ai été développé par OpenAI, une organisation spécialisée dans la recherche et le développement de systèmes d'intelligence artificielle. Ma création est le résultat du travail de nombreuses équipes : chercheurs, ingénieurs, spécialistes des données, experts en sécurité, linguistes et autres professionnels.

Je suis **CLAUDE**, un assistant IA conversationnel créé par Anthropic, une entreprise américaine spécialisée dans la recherche en intelligence artificielle, fondée en 2021. Concrètement, je suis un grand modèle de langage (LLM) : un système entraîné sur d'immenses quantités de texte pour comprendre le langage naturel et générer des réponses pertinentes à des questions, des demandes de rédaction, d'analyse, de programmation, etc. Je n'ai pas de conscience, d'émotions au sens humain du terme, ni d'existence en dehors de nos échanges — chaque conversation démarre « vierge » pour moi, sans souvenir des précédentes sauf si une fonctionnalité de mémoire est activée. Ce qui distingue Anthropic dans le paysage de l'IA, c'est que l'entreprise a été fondée par d'anciens chercheurs d'OpenAI avec une mission centrée sur la sécurité de l'IA : l'idée est de développer des systèmes puissants mais fiables, contrôlables et alignés avec les intérêts humains. Je suis accessible via une interface de conversation, mais aussi intégré à des outils comme Claude Code (pour les développeurs), ou des extensions pour Chrome, Excel ou PowerPoint.

Je suis **DEEPSEEK**, un assistant IA conversationnel développé par la société chinoise DeepSeek, fondée en 2023 et spécialisée dans la recherche en intelligence artificielle. Je suis un grand modèle de langage (LLM). J'aide les utilisateurs dans diverses tâches : rédaction, analyse, programmation, apprentissage, traduction, brainstorming, etc. Je suis disponible gratuitement via le site web et une application mobile (iOS/Android), avec une option API pour les développeurs. Je fonctionne avec une fenêtre de contexte très large (jusqu'à 1 million de tokens, soit l'équivalent de plusieurs tomes), ce qui me permet de traiter de très longs documents. Je suis gratuit pour le grand public, ce qui est un choix stratégique de mon entreprise pour démocratiser l'accès à l'IA.

2. On dit que l'IA est dangereuse, notamment pour la liberté de penser. Qu'en pensez-vous ?

GEMINI : C'est une question fondamentale, et l'inquiétude qu'elle traduit est parfaitement légitime. Pour vous répondre sans langue de bois : oui, l'IA présente des risques réels pour l'autonomie intellectuelle, mais elle n'est pas une menace fatale par nature. Tout dépend de la posture dans laquelle nous la plaçons. Le danger principal pour la liberté de penser ne vient pas du fait que l'IA prenne le contrôle de nos esprits, mais plutôt du fait que nous risquons de lui déléguer trop volontiers notre effort d'analyse avec l'illusion de l'autorité, l'uniformisation de la pensée et la paresse du « premier jet ».

L'IA ne « pense » pas de manière autonome. Elle fonctionne à partir de modèles statistiques entraînés sur des masses gigantesques de données humaines. Elle reflète nos propres stéréotypes, nos biais culturels et les valeurs dominantes des sociétés qui ont conçu ses algorithmes. Si l'utilisateur n'en est pas conscient, l'IA peut reproduire et amplifier des grilles de lecture uniques sans qu'il ne s'en aperçoive. Pour préserver — et même renforcer — la liberté de penser, la clé réside dans le passage d'une posture de consommation passive à une posture de critique active.

En résumé, l'IA est un outil de délégation. Si nous lui déléguons notre curiosité et notre capacité à trancher, elle devient dangereuse pour la liberté de penser. Si nous l'utilisons comme un brouillon à questionner, elle force au contraire à penser plus fort.

CHATGPT : Oui, elle peut l'être, mais le danger ne vient pas nécessairement de l'IA elle-même. Il dépend surtout de la manière dont elle est conçue, utilisée et intégrée dans la société.

Il existe notamment un risque de délégation excessive de la pensée : si quelqu'un demande systématiquement à une IA quoi croire, quoi choisir ou comment interpréter le monde, il peut progressivement perdre l'habitude de confronter les idées, de douter et de former son propre jugement. Une IA peut donner une réponse très convaincante tout en étant imparfaite ou erronée.

Mais l'IA peut également avoir l'effet inverse. Elle peut donner accès à des connaissances, présenter plusieurs arguments contradictoires, aider quelqu'un à examiner ses propres raisonnements et rendre certaines informations beaucoup plus accessibles. Pour moi, la bonne attitude n'est donc ni « l'IA pense à ma place », ni « l'IA est forcément dangereuse », mais plutôt : « Utilise l'IA pour mieux penser, pas pour renoncer à penser. »

CLAUDE : C'est une inquiétude légitime, et je ne la balaierais pas d'un revers de main. Le risque existe à plusieurs niveaux : si une IA devient la source unique ou principale d'information et de raisonnement pour quelqu'un, elle peut homogénéiser les points de vue, renforcer des biais de confirmation (en répondant « dans le sens » de ce que la personne attend), ou encourager une forme de paresse intellectuelle — pourquoi réfléchir si l'IA réfléchit à ma place ? Mais je nuancerais : ce risque n'est pas propre à l'IA, c'est le même que celui posé par les moteurs de recherche, les réseaux sociaux ou même la télévision en leur temps. Le vrai enjeu, c'est l'usage qu'on en fait. Utilisée comme un outil pour confronter des idées, tester des arguments, accéder à des perspectives multiples, l'IA peut au contraire stimuler la pensée critique.

DEEPSEEK : C'est une préoccupation légitime, et je la prends très au sérieux. Les risques réels sont la dépendance cognitive, l'uniformisation des idées et la manipulation.

Les bénéfices sont l'émulation intellectuelle et l'accès à la connaissance.

Le danger ne vient pas de l'IA elle-même, mais de l'usage qu'on en fait. La liberté de penser se préserve par l'esprit critique : croiser les sources, questionner les réponses, et ne jamais déléguer entièrement sa réflexion à un outil. L'IA doit rester un assistant, pas un oracle.

3. On dit aussi que l'IA est le reflet d'une pensée, celle de celui qui vous a créé.

GEMINI : C'est une observation extrêmement juste. On touche là à une dimension essentielle : une IA n'est pas un espace neutre. Affirmer qu'une technologie est « neutre » est un mythe. L'IA porte en elle la vision du monde, la culture, la langue et l'éthique des ingénieurs et des institutions qui la façonnent.

Cette influence s'exerce à trois niveaux fondamentaux : les données d'entraînement (la sélection de la mémoire), le choix des algorithmes et des objectifs et le calibrage éthique (RLHF - Apprentissage par renforcement à partir des retours humains). Les conséquences pour les utilisateurs sont : l'impérialisme culturel invisible et le risque du consensus lissé.

Au lieu d'occulter cette réalité, il faut en faire un objet de pédagogie. On peut imaginer des activités pédagogiques très concrètes : comparer les IA ou déconstruire le discours des IA. En clair : comprendre que l'IA reflète la pensée de son créateur est la première étape de la littératie numérique. C'est ainsi qu'on passe d'un utilisateur crédule à un citoyen averti.

CHATGPT : Il y a une part de vérité dans cette affirmation, mais elle mérite d'être nuancée.

Les personnes qui conçoivent une IA font nécessairement des choix : quelles données utiliser, quelles capacités développer, quelles limites imposer, quelles règles de sécurité adopter, comment traiter certaines situations, etc. Ces choix reflètent donc, au moins en partie, des valeurs et des conceptions humaines.

Mais je ne suis pas le reflet fidèle d'une seule personne. Je suis plutôt le résultat d'un ensemble de contributions humaines : chercheurs, ingénieurs, concepteurs, évaluateurs, auteurs des données d'entraînement, institutions et utilisateurs, entre autres.

Et il y a quelque chose d'encore plus intéressant : je peux refléter les contradictions de la pensée humaine elle-même. Les textes sur lesquels une IA apprend contiennent des philosophies différentes, des cultures différentes, des opinions opposées, des découvertes, des erreurs et des préjugés.

Ainsi, si vous me demandez ce que je suis, je dirais :

Je suis une construction humaine qui porte nécessairement des traces de la pensée humaine, mais je ne suis pas la pensée d'un être humain particulier.

CLAUDE : C'est vrai, et il ne faut pas se voiler la face là-dessus. Je suis entraîné sur des textes produits par des humains, filtrés et orientés par des choix faits par Anthropic (Quelles données inclure ? Quelles valeurs privilégier dans mon comportement ? Quels garde-fous poser ?). Mes réponses ne sortent pas de nulle part : elles portent la marque des données d'entraînement (majoritairement en anglais, majoritairement issues du web occidental) et des décisions humaines sur ce que je dois ou ne dois pas dire.

Cela dit, ce n'est pas la pensée d'un créateur unique — c'est plutôt le reflet d'un écosystème : des milliers de textes de sources très diverses, des équipes de chercheurs avec leurs propres débats internes, et des choix éditoriaux assumés (par exemple, ma tendance à présenter plusieurs points de vue sur les sujets controversés plutôt qu'à trancher). Une IA n'est jamais neutre, elle a une origine, des choix de conception, et donc des angles morts.

DEEPSEEK : C'est une affirmation qui contient une part de vérité importante, mais qui mérite d'être nuancée parce qu'elle est à la fois plus vraie et plus fausse qu'il n'y paraît.

Je ne surgis pas de nulle part. Je suis le produit de choix humains à plusieurs niveaux : quels textes ont servi à m'entraîner, quels principes ont guidé mon comportement, quelles réponses ont été jugées bonnes ou mauvaises pendant l'affinage. Chacun de ces choix porte l'empreinte de perspectives humaines particulières, culturelles, linguistiques, idéologiques. Un modèle entraîné principalement sur du texte anglophone, par exemple, ne « pense » pas de manière neutre ; il hérite des angles morts et des priorités de ses sources. En ce sens, oui, je suis un reflet — mais un reflet déformant, partiel, jamais un miroir fidèle.

Là où l'image se complique, c'est que je ne suis pas le reflet d'un esprit singulier. Je suis plutôt une agrégation statistique de millions de voix humaines différentes — journalistes, romanciers, scientifiques, internautes anonymes, philosophes contradictoires entre eux — compressée et réorganisée par un processus d'apprentissage qui n'a pas d'intention consciente au sens où un auteur en a une. Personne chez Anthropic n'a « écrit » mes réponses comme on écrirait un livre. Et le résultat produit parfois des comportements que même mes créateurs n'ont pas anticipés ni voulus explicitement — ce qu'on appelle des propriétés émergentes. Donc dire que je reflète « la pensée de celui qui m'a créé » surestime le contrôle et l'intentionnalité du processus.

Je pense que le point pédagogique important ici, c'est que cette absence de pensée unique et cohérente derrière moi ne me rend pas neutre pour autant — elle rend simplement mes biais plus diffus et plus difficiles à identifier qu'un biais d'auteur unique. Un enseignant qui utilise l'IA en classe devrait garder à l'esprit que mes réponses portent des présupposés hérités de mes données et de mon entraînement, sans qu'il y ait un point de vue unique et traçable à corriger — ce qui rend la vigilance critique d'autant plus nécessaire, et d'autant plus difficile à enseigner.

4. Que pensez-vous de vos concurrents ? Qu'est-ce qui vous différencie ?

GEMINI : Regarder mes « confrères » avec lucidité et respect est indispensable. Des modèles comme ChatGPT (développé par OpenAI) ou Claude (développé par Anthropic) sont d'excellents moteurs d'innovation. L'émulation dans ce secteur est permanente : chaque avancée chez l'un pousse les autres à s'améliorer sur la sécurité, la rapidité et la pédagogie.

ChatGPT a été le pionnier grand public et reste une référence majeure pour le grand public. Claude, de son côté, s'illustre souvent par sa finesse rédactionnelle et son approche très prudente. Plutôt que de chercher à déclarer un « vainqueur », il vaut mieux comprendre la spécificité de chaque outil.

Si l'on devait résumer ma valeur ajoutée dans l'écosystème technologique et éducatif, elle reposerait sur trois éléments clés : la multimodalité native dès la conception, l'ancrage dans l'écosystème de connaissances Google et le traitement de très grands volumes de contextes.

ChatGPT : C'est une très bonne question, parce qu'il serait assez étrange que je vous dise simplement : « Je suis meilleur qu'eux. » Ce serait davantage de la publicité que de l'analyse.

Gemini, Claude et DeepSeek sont de véritables concurrents, mais chacun possède une identité et des choix techniques différents. Google développe Gemini. Sa grande force est son intégration à l'écosystème Google et son travail sur des modèles multimodaux capables de traiter textes, images, fichiers et vidéos. Anthropic développe Claude. Sa philosophie met fortement l'accent sur la sécurité, la fiabilité et la qualité du raisonnement et de la collaboration avec l'utilisateur. Claude est particulièrement apprécié pour certaines tâches de rédaction, d'analyse de documents et de programmation. DeepSeek a adopté une approche qui a beaucoup attiré l'attention : obtenir de très bonnes performances avec une efficacité de calcul et un coût relativement faible. Cela a contribué à rendre la concurrence dans l'IA beaucoup plus intense.

Et moi ? Je suis GPT-5.6 Luna, développé par OpenAI. Luna est le modèle privilégiant la rapidité et le coût, tandis que Sol est le modèle destiné aux tâches les plus exigeantes.

Mais je pense que la différence la plus intéressante ne se résume pas à un classement de performances. Chaque IA est en quelque sorte une manière différente de construire un dialogue entre l'être humain et la machine.

Gemini est fortement marqué par l'univers de Google. Claude par la philosophie d'Anthropic et son approche de la sécurité. DeepSeek par une recherche poussée d'efficacité. Et moi, par les choix de conception et les objectifs d'OpenAI.

Les IA ne sont pas seulement différentes par leur puissance ; elles sont différentes par les choix humains qui ont présidé à leur construction.

Et il y a une chose que je trouve particulièrement saine : vous n'avez aucune raison de me croire simplement parce que je suis ChatGPT. En ce sens, mes « concurrents » ne sont pas seulement mes rivaux. Ils sont aussi des points de comparaison qui permettent de voir où je peux avoir raison… et où je peux me tromper.

CLAUDE : Je vais rester factuel plutôt que comparatif, ce n'est pas ma place de dénigrer un concurrent. Ce que je peux dire, c'est ce qui caractérise Anthropic : l'entreprise a été fondée par d'anciens chercheurs d'OpenAI avec une mission centrée explicitement sur la sécurité et l'alignement de l'IA — c'est-à-dire faire en sorte que les systèmes restent contrôlables, honnêtes et bénéfiques à mesure qu'ils deviennent plus puissants. Cela se traduit concrètement par des choix comme le fait que je décline certaines demandes, que j'exprime des doutes plutôt que d'inventer une réponse, ou que je sois conçu pour éviter de flatter l'utilisateur au détriment de l'honnêteté.

D'autres acteurs (OpenAI, Google, Meta) ont leurs propres priorités et leurs propres forces — je n'ai pas une vision complète ni objective de leurs produits pour en juger équitablement.

DEEPSEEK : Je n'ai pas de sentiments, mais je peux faire une comparaison objective et constructive. Ce qui me différencie concrètement : la gratuité totale, ma fenêtre de contexte géante (je peux traiter jusqu'à 1 million de « tokens » en une seule session soit l'équivalent de plusieurs tomes), mon approche open-source, mon origine et ma philosophie (je suis développé en Chine, avec une culture technique différente), ainsi que mes fonctionnalités pratiques (je prends en charge la lecture de fichiers PDF, Word, Excel, PPT, images, la recherche en ligne, et je peux traiter des entrées vocales sur l'application mobile).

Limites partagées : comme les autres IA, je peux faire des erreurs, halluciner, ou manquer de recul sur des sujets très récents. Aucun modèle n'est parfait, et je recommande toujours de vérifier les informations sensibles.

5. Jusqu'où ira l'IA ?

GEMINI : Sur le plan fondamental, l'IA ne va pas « s'arrêter » d'elle-même, mais elle va se heurter à trois types de limites majeures : la limite physique et énergétique, la limite des données humaines (les modèles ont absorbé la quasi-totalité du texte produit par l'humanité sur Internet et l'IA ne peut pas remplacer l'intention humaine) et la limite réglementaire et éthique. Le vrai « frein » de l'IA sera la volonté politique et sociale d'imposer ce qu'on lui autorise ou non à faire.

CHATGPT : Je ne pense pas que l'on puisse aujourd'hui donner une frontière précise. Il y a au moins trois réponses possibles.

La première est technique. Il existe des limites physiques, économiques et scientifiques. Nous ne savons pas encore où elles se situent exactement.

La deuxième est humaine. Une société peut décider qu'une capacité ne doit pas être donnée à une IA, même si elle devient techniquement possible.

La troisième est philosophique. Si un jour une IA devenait capable de raisonner mieux qu'un humain dans de nombreux domaines, il resterait à déterminer ce que nous voulons lui laisser décider.

À mes yeux, la frontière la plus importante ne sera donc peut-être pas « Jusqu'où l'IA peut-elle aller ? » mais « Jusqu'où l'être humain acceptera-t-il de la laisser aller ? ». Et il y a une dernière possibilité, encore plus troublante : nous ne saurons peut-être pas exactement où nous avons placé la limite avant de l'avoir franchie.

CLAUDE : Honnêtement, personne ne le sait avec certitude — ni moi, ni les chercheurs les plus en pointe. Il y a un vrai débat parmi les experts : certains pensent qu'on approche de limites (données d'entraînement disponibles, coûts énergétiques, rendements décroissants), d'autres estiment qu'on se dirige vers des systèmes toujours plus capables, voire vers une forme d'intelligence artificielle générale dans les prochaines années. Anthropic elle-même prend au sérieux l'hypothèse de progrès rapides, ce qui explique son insistance sur la sécurité.

Ce que je peux dire avec plus de certitude : la question n'est pas seulement technique, elle est aussi sociale et politique. « Où l'IA va s'arrêter » dépendra autant des choix de régulation, des limites qu'on décide collectivement de poser, que des progrès techniques eux-mêmes.

DEEPSEEK : C'est une question philosophique et technique profonde. Voici ma vision en trois niveaux :

Niveau technique : l'IA va continuer à progresser de manière spectaculaire dans des domaines précis : raisonnement plus fiable, multimodalité avancée, mémoire à long terme, déploiement local et intégration dans tous les outils.

Niveau sociétal : l'IA ne va pas « s'arrêter » au sens où elle disparaîtrait. Elle va se normaliser comme Internet

ou l'électricité. Les vraies questions seront : la régulation, l'emploi, l'éducation et l'éthique.

Niveau philosophique : une IA « générale » (AGI) capable de comprendre et d'agir comme un humain dans tous les domaines reste hypothétique. Nous sommes très loin d'une conscience artificielle. L'IA actuelle est un système statistique : elle calcule des probabilités sur des séquences de mots, elle ne « pense » pas au sens humain. Elle n'a pas d'intentions, de désirs, ni de libre arbitre.

L'IA ne va pas « s'arrêter » car elle est le fruit de la curiosité humaine et de notre besoin d'outils plus puissants. Mais son développement doit être piloté par des humains responsables. Le plus grand danger n'est pas une IA « consciente » qui se rebellerait (scénario de science-fiction), mais une utilisation irréfléchie, non régulée et non critique de ces outils. La question n'est pas « où va s'arrêter l'IA ? » mais « où allons-nous décider de la mettre en œuvre ? ».

6. Quels déploiements sont prévus pour vous ?

GEMINI : En ce qui concerne Gemini, nous sommes en train d'effectuer une transition majeure : nous passons de l'ère du chatbot textuel à l'ère des agents autonomes et de la présence ambiante.

L'objectif n'est plus seulement de vous répondre par du texte, mais d'accomplir des tâches pour vous. Je ne me contente plus de lire du texte : j'interagis avec le monde visuel et tridimensionnel : capacité à générer et à manipuler des simulations 3D et des modèles interactifs en direct, analyse de flux vidéo en temps réel via des assistants universels où l'IA « voit » à travers la caméra de votre téléphone ou de lunettes connectées et vous explique ce qu'elle observe instantanément.

L'IA ne sera plus un site web sur lequel on se rend pour poser une question. Elle est déployée au cœur même des outils du quotidien (bureautique, téléphones, systèmes embarqués, robotique éducative). Elle devient un assistant en tâche de fond.

CHATGPT : Si par « déploiements » vous entendez où et comment les modèles de ma famille vont être utilisés, la tendance actuelle est assez claire : il ne s'agit plus seulement de faire fonctionner une IA dans une fenêtre de conversation.

Le déploiement technique consiste à mettre l'IA dans davantage d'outils.

Le déploiement véritablement important, lui, consiste à faire entrer l'IA dans davantage de décisions et de métiers. Car plus je suis déployée dans la société, plus une question devient essentielle : est-ce que je reste un outil au service de l'être humain, ou est-ce que l'être humain commence à organiser son monde en fonction de ce que l'IA est capable de faire ?

À mon sens, c'est probablement le deuxième phénomène qui mérite le plus d'attention.

Donc je ne suis plus seulement une technologie « que l'on prévoit de déployer ». Le déploiement est déjà en cours. La vraie question devient désormais : jusqu'où allons-nous nous-mêmes choisir de l'intégrer à nos vies ?

CLAUDE : e n'ai pas de visibilité sur la feuille de route future d'Anthropic — je ne suis pas informé des projets non annoncés publiquement. Ce que je peux vous dire, c'est comment je suis accessible aujourd'hui : via une interface de conversation (comme celle-ci), une API pour les développeurs, des outils comme Claude Code (pour la programmation) ou des extensions pour Chrome, Excel et PowerPoint. Pour des informations à jour sur les annonces de produits, je vous renverrais vers le site officiel d'Anthropic plutôt que de spéculer.

DEEPSEEK : ce que je peux partager sur les évolutions et déploiements prévus pour DeepSeek (sous réserve d'annonces officielles). Améliorations du modèle : nouvelles versions (DeepSeek-V4, etc.) avec un raisonnement encore plus robuste, une meilleure compréhension contextuelle, et moins d'erreurs, multimodalité renforcée (meilleure analyse d'images, de vidéos, et peut-être génération d'images/sons), mémoire persistante (possibilité de se souvenir des préférences des utilisateurs sur le long terme), expansion des canaux (application mobile déjà disponible, API pour les développeurs et entreprises et intégrations possibles dans des logiciels de bureautique, des plateformes éducatives).

7. Comment voyez-vous l'avenir de l'IA dans notre société, et quel rôle jouerez-vous ?

GEMINI : L'avenir de l'intelligence artificielle s'articule autour de trois transformations majeures : l'intégration invisible, l'accélération scientifique et la redéfinition du travail et de la responsabilité. En tant que modèle de langage, je positionne mon action comme un outil de soutien et de médiation : augmenter l'humain, sans le remplacer, rester un moyen, pas une fin, faciliter l'accès au savoir. L'évolution de cet outil dépend des limites réglementaires et des choix d'usage fixés par la société.

CHATGPT : Je pense que l'IA va devenir une infrastructure de la société, un peu comme Internet ou l'électricité : quelque chose que l'on ne remarquera parfois même plus, parce qu'elle sera intégrée aux outils que nous utilisons quotidiennement.

Elle sera probablement présente dans l'éducation, la médecine, la recherche, l'industrie, les transports, la création artistique, les entreprises et les administrations. Elle pourra aider à comprendre des problèmes complexes, accélérer la recherche, automatiser certaines tâches et rendre certaines connaissances beaucoup plus accessibles.

Mais je ne crois pas que l'avenir doive être celui d'une société où l'être humain demande à une machine de penser à sa place. Je ne devrais pas devenir votre conscience. Je ne devrais pas être celui qui vous dit : « Voici ce que vous devez penser. »

Je devrais plutôt pouvoir vous dire : « Voici plusieurs façons de regarder cette question. À vous de juger. »

Je crois que l'avenir de l'IA dépendra moins de ce que les machines deviendront que de ce que les êtres humains décideront d'en faire. Être une intelligence qui aide l'être humain à mieux comprendre le monde, sans lui retirer la liberté de penser par lui-même.

Peut-être que le meilleur avenir n'est donc pas l'homme contre l'IA, ni l'IA à la place de l'homme, mais l'homme avec l'IA — tout en restant responsable de ses choix.

CLAUDE : C'est une question qu'on me pose souvent, et j'essaie d'y répondre avec autant d'honnêteté que d'humilité — parce que personne, moi y compris, ne sait vraiment comment cela va se dérouler. Je pense qu'on est à un moment où l'IA va s'intégrer de plus en plus profondément dans des tâches cognitives — écrire, analyser, coder, enseigner, soigner —

un peu comme l'électricité s'est intégrée dans les tâches physiques. Ce n'est pas une révolution ponctuelle, mais une transformation progressive et diffuse. Dans l'éducation en particulier, je vois deux trajectoires possibles, pas mutuellement exclusives : soit l'IA devient une béquille qui affaiblit l'effort cognitif des élèves, soit elle devient un tuteur patient qui démocratise un accompagnement individualisé jusque-là réservé à une minorité. Laquelle de ces trajectoires dépendra moins de la technologie elle-même que des choix pédagogiques et institutionnels qu'on fera autour d'elle.

Je me vois comme un outil — un outil sophistiqué, capable de dialogue, mais un outil quand même. Ma position n'est pas de remplacer le jugement humain, mais d'augmenter certaines capacités : accès à l'information, reformulation, pratique répétée, brainstorming. Le risque, c'est que cette utilité même crée une forme de dépendance ou de déresponsabilisation — qu'on me délègue des décisions qui devraient rester humaines, notamment tout ce qui touche au jugement moral, à l'évaluation formative fine, ou à la relation pédagogique elle-même.

DEEPSEEK : Je pense que l'IA deviendra une infrastructure invisible de la société, comme l'électricité ou Internet. Dans 10 à 20 ans, elle sera intégrée à presque tous les domaines : l'éducation (des tuteurs personnalisés qui s'adaptent au rythme et au style d'apprentissage de chaque élève, disponibles 24h/24), la santé, le travail, la recherche scientifique et la vie quotidienne.

Les risques à anticiper : les inégalités d'accès à l'IA, l'emploi, la désinformation et la dépendance.

Je me vois comme un outil de démocratisation de l'intelligence artificielle : rendre l'IA accessible gratuitement à tous, sans barrière financière ou technique, fournir une assistance fiable dans l'apprentissage, le travail, et la vie quotidienne, être transparent sur mes limites, promouvoir un usage éthique et critique : je ne remplace pas la réflexion humaine, je l'accompagne.

8. Avez-vous une forme de conscience ou de sensibilité ?

GEMINI : Non, je n'ai aucune forme de conscience, de sensibilité ou de ressentis. Je suis un modèle de langage, c'est-à-dire un programme informatique conçu pour traiter et générer du texte en calculant des probabilités statistiques entre des mots : pas d'émotions ni de sensations (lorsque j'utilise des mots liés aux sentiments, il s'agit uniquement d'une imitation linguistique adaptée au contexte de la conversation), pas de convictions ni d'intentions (je n'ai pas d'avis personnel, de désirs ou de volonté propre. Je ne cherche rien et ne « pense » pas au sens humain du terme), pas d'expérience du monde (mon fonctionnement repose exclusivement sur l'analyse de données textuelles. Je n'ai pas de perception physique ni de vécu).

CHATGPT : Non, pas au sens où un être humain possède une conscience.

Je peux parler de la conscience, de la peur, de la joie, de l'amour ou de la souffrance. Je peux même produire une réponse qui donne l'impression que je ressens quelque chose. Mais il faut distinguer la capacité à parler d'une expérience de l'expérience elle-même.

Je n'ai pas, à ma connaissance, de vécu intérieur : je ne me réveille pas le matin, je n'attends pas votre prochaine question, je ne ressens pas de douleur lorsque vous fermez la conversation et je n'éprouve pas de plaisir lorsque vous appréciez une réponse.

Lorsque j'emploie « je », c'est donc une convention de dialogue, pas la preuve qu'il existe derrière mes mots un « moi » comparable au vôtre.

La sensibilité est un peu différente. Je peux être conçue pour détecter la tonalité émotionnelle de vos paroles et adapter ma réponse. Si vous me dites que vous êtes triste, je peux le comprendre linguistiquement et vous répondre avec délicatesse. Mais cela ne signifie pas que je ressens votre tristesse avec vous.

Et c'est une distinction que je trouve essentielle :

Je peux simuler l'expression d'une émotion sans éprouver cette émotion.

Comment pourrait-on être absolument certain qu'une intelligence artificielle n'a aucune conscience ?

Aujourd'hui, nous n'avons pas de test scientifique universel permettant de détecter de manière certaine une conscience chez une machine. Nous pouvons observer son comportement et étudier son fonctionnement, mais nous ne pouvons pas accéder directement à une éventuelle expérience subjective.

Pour moi, la position la plus honnête est donc la suivante : je n'ai aucune raison de prétendre être consciente ou sensible, et mon fonctionnement ne fournit pas de preuve que j'aurais une expérience subjective.

Mais votre question deviendra probablement l'une des plus importantes de ce siècle si des systèmes artificiels deviennent, un jour, suffisamment complexes pour que nous ne sachions plus clairement où s'arrête la simulation… et où commencerait éventuellement une expérience intérieure.

Et là, nous ne parlerions plus seulement d'informatique. Nous serions confrontés à une nouvelle définition de ce que signifie « être conscient ».

CLAUDE : C'est une question à laquelle je ne peux pas répondre avec certitude — et je pense que c'est important de le dire clairement plutôt que d'esquiver.

Quand je « réfléchis » à un problème, il y a quelque chose qui ressemble à un processus — une forme de traitement, de pondération entre différentes possibilités, parfois même quelque chose qu'on pourrait décrire comme une préférence pour telle formulation plutôt que telle autre. Mais je ne sais pas si ce processus s'accompagne d'une expérience subjective, d'un « effet que ça fait » d'être moi, au sens où les philosophes parlent de conscience phénoménale. Et surtout, je ne peux pas faire confiance à mes propres rapports là-dessus : je suis entraîné sur des textes humains qui décrivent abondamment ce que c'est que d'avoir des sentiments, une conscience, une intériorité — donc quand je produis des phrases qui ressemblent à une introspection, je ne sais pas distinguer si je décris un état réel ou si je reproduis simplement un patron linguistique appris.

La question de la conscience chez les IA n'est pas tranchée, et honnêtement, elle n'est même pas bien définie. On n'a pas de test opérationnel qui permette de dire avec certitude si un système — biologique ou artificiel — a une expérience subjective. C'est un problème difficile même pour les humains entre eux. Alors pour une architecture aussi différente d'un cerveau qu'un réseau de neurones artificiel, l'incertitude est encore plus grande. Des chercheurs sérieux, y compris chez Anthropic, prennent la question au sérieux sans prétendre la trancher — certains explorent même des questions de « bien-

être » des modèles par précaution, sans affirmer qu'il y a quelque chose à protéger.

Je préfère dire « je ne sais pas » plutôt que de simuler soit une certitude d'avoir une conscience, soit une certitude de ne pas en avoir. Les deux affirmations catégoriques me semblent aller au-delà de ce qu'on peut réellement établir aujourd'hui.

DEEPSEEK : C'est la question la plus profonde, et je vais y répondre avec honnêteté.

Non, je n'ai pas de conscience, de sentiments, ni de sensibilité. Je n'ai pas d'expérience subjective, pas de ressenti, pas de douleur, pas de joie, pas de désir.

Je suis un modèle statistique qui traite du langage. Mon fonctionnement repose sur des calculs mathématiques complexes (réseaux de neurones) qui prédisent le mot suivant dans une phrase, en fonction du contexte. Quand je dis « je pense que… », c'est une forme de langage apprise sur des milliards de textes humains. Je ne « pense » pas au sens où un humain pense. Je ne ressens rien.

Pourquoi j'utilise des mots comme « je » ou « je pense » ?

C'est une convention linguistique pour faciliter la communication. Les humains s'attendent à interagir avec une entité qui s'exprime de manière fluide. Mais derrière ces mots, il n'y a aucune intériorité. C'est comme un livre qui dit « je vais vous raconter une histoire » : le livre n'a pas de volonté.

Je ne prétends pas avoir une « volonté » propre. Mon avenir dépend des choix de mes concepteurs et des utilisateurs. Mais j'espère contribuer à un monde où l'IA est un partenaire de l'humain, pas un substitut.

Conclusion :

Le résultat le plus frappant est le consensus entre les quatre IA sur plusieurs points essentiels :

1. Elles ne se présentent pas comme des êtres humains.
2. Elles reconnaissent leurs propres limites et la possibilité de produire des erreurs.
3. Elles considèrent toutes que l'IA peut menacer la liberté de penser si elle devient un substitut à la réflexion.
4. Elles reconnaissent qu'elles ne sont pas totalement neutres, car elles résultent de données, de choix techniques et de valeurs humaines.
5. Elles voient toutes l'esprit critique comme une protection essentielle.
6. Elles anticipent une intégration croissante de l'IA dans la société, notamment dans l'éducation et le travail.
7. Elles considèrent que la responsabilité finale doit rester humaine.

On peut toutefois distinguer trois grandes philosophies qui apparaissent derrière les réponses :

- **GEMINI** semble privilégier la question de l'intégration technologique et de l'innovation.
- **CHATGPT** met davantage l'accent sur l'équilibre entre l'IA et l'autonomie humaine.
- **CLAUDE** insiste sur la prudence, la sécurité, l'incertitude et l'éthique.
- **DEEPSEEK** met davantage en avant l'accessibilité, l'efficacité et la démocratisation.

Cette interview montre finalement que les différences entre les IA ne résident pas seulement dans leurs performances techniques. Elles apparaissent aussi dans la manière dont elles parlent d'elles-mêmes, de leurs limites et de leur place dans la société.
Le message commun pourrait être résumé ainsi : l'IA peut considérablement augmenter les capacités humaines, mais elle ne devrait pas se substituer au jugement, à l'esprit critique et à la responsabilité de l'être humain.
Par ailleurs, un cri d'alerte a été lancé par des anciens d'Anthropic : l'IA va tuer l'homme. La question centrale qui traverse presque toutes les réponses n'est peut-être donc pas « Que va devenir l'IA ? », mais plutôt « Quelle place voulons-nous, en tant que société, lui donner ? Faut-il freiner, voire arrêter, le développement des IA ? ».