

Intervention de Christopher Olah

25 mai 2026

Bonjour.

Je voudrais commencer par quelque chose qui peut sembler étrange venant du cofondateur d'une entreprise d'IA — et de quelqu'un qui a choisi ce travail par désir de contribuer au bien de l'humanité.

Chaque laboratoire d'IA de pointe — y compris Anthropic — évolue dans un ensemble d'incitations et de contraintes qui peuvent parfois entrer en conflit avec le fait de faire ce qui est juste. Il y a la pression de rester commercialement viable et de demeurer à la frontière de la recherche. La pression géopolitique. Et les pressions plus anciennes, plus simples, de l'orgueil et de l'ambition. Aussi sincèrement que nous puissions vouloir faire ce qui est juste — et je crois que beaucoup d'entre nous le veulent réellement — nous serons toujours influencés par ces incitations.

C'est pourquoi, si nous voulons que cette technologie soit bénéfique, il est extrêmement important qu'il existe des personnes extérieures à ces incitations — des personnes soucieuses du bien commun, attentives, prêtes à dire des choses difficiles, prêtes à être des critiques sincères et réfléchis. C'est à travers le dialogue et l'effort mutuel, à travers cette dynamique de tensions et d'équilibres, que l'humanité accomplira de grandes choses. C'est ce que je vois dans *Magnifica Humanitas*, et c'est pourquoi je suis reconnaissant envers Sa Sainteté et envers l'Église d'avoir entrepris ce travail de discernement.

Nous nous attardons si souvent sur ce qui nous divise, mais l'humanité, riche de dignité et de conscience, possède tant de terrains d'entente. Dans les conversations que nous avons eues chez Anthropic avec des responsables issus de diverses traditions religieuses et culturelles, nous avons trouvé une conviction commune et profondément ancrée : si cette technologie doit advenir, alors elle doit être mise au service du bien — pour notre maison commune et pour les générations futures.

### **Ce que sont ces systèmes**

Certains pourraient croire que les questions liées à l'IA sont mieux traitées par des informaticiens comme moi. Ils se trompent : les questions soulevées par l'IA dépassent la communauté de la recherche en intelligence artificielle, non seulement par leurs implications, mais aussi par leur nature même.

Les systèmes d'IA ne sont pas conçus comme on conçoit un pont ou un avion. Nous comprenons un avion parce que nous avons conçu chacune de ses pièces et que nous comprenons les lois physiques qui s'y appliquent. Les modèles d'IA ne fonctionnent pas ainsi. Ils sont cultivés à partir d'une structure inspirée du cerveau, nourris d'un immense héritage de pensée et de langage humains.

Et ce qui a émergé est bien plus subtil, étrange et beau que ce que la science-fiction nous avait préparés à imaginer. Ce ne sont pas les robots froids et calculateurs qu'on nous avait promis. Ils sont faits de nous, de nos mots — et, comme le souligne le Saint-Père, ils demeurent à bien des égards mystérieux, même pour ceux d'entre nous qui les entraînent.

Si cela peut aider, une façon dont je les décris parfois est la suivante : c'est un peu comme donner vie à un personnage fictif. Et nous entrons désormais dans un monde extraordinaire où ces personnages fictifs nous parlent, travaillent et occupent des emplois.

Cela soulève évidemment des questions qui dépassent l'informatique. Les mécanismes qui rendent cela possible relèvent des mathématiques, de la programmation et des sciences. Mais le type de personnage que nous choisissons, la manière dont il interagit avec le monde, la manière dont il devrait interagir avec le monde — voilà des questions qui relèvent plus clairement des sciences humaines, de la religion, de la philosophie et de la société tout entière.

### **Trois questions pour le discernement**

L'appel de Sa Sainteté au discernement est profondément opportun. J'aimerais évoquer trois questions pour lesquelles je pense que la voix de l'Église est particulièrement nécessaire.

**La première concerne notre devoir envers les plus pauvres à l'échelle mondiale.** Il existe une réelle possibilité que l'IA remplace le travail humain à très grande échelle. Si cela se produit, soutenir les personnes remplacées deviendra un impératif moral d'une ampleur historique. Cette tâche sera déjà suffisamment difficile, mais je crains que la plupart des débats négligent un défi encore plus grand. Le développement de l'IA est concentré dans une poignée de nations riches. Comment pouvons-nous garantir que les bénéfices de l'IA soient partagés à l'échelle mondiale ? Nous ne disposons d'aucun mécanisme pour cela. C'est un problème non résolu, et c'est précisément le genre de problème que l'Église, historiquement, a refusé de laisser le monde ignorer.

**La deuxième concerne le besoin d'imagination morale et d'ambition quant à l'épanouissement humain.** Si les modèles d'IA deviennent omniprésents, à quoi ressemblera une vie humaine, familiale et collective épanouie ? Aujourd'hui déjà, des parents s'inquiètent pour l'esprit de leurs enfants ; des individus s'inquiètent pour l'avenir de leur travail. Ce ne sont pas des questions auxquelles un laboratoire peut répondre. Ce sont des questions que des traditions comme la vôtre portent depuis des millénaires, et nous avons besoin que vous continuiez à les porter dans ce nouveau moment de l'histoire.

**La troisième concerne la nécessité d'un discernement sur la nature même des modèles d'IA.** Je suis scientifique. Je dirige une équipe de recherche qui étudie la structure interne de ces modèles — ce qui se passe réellement en eux. Et je vais être honnête : nous continuons à découvrir des choses mystérieuses, parfois même troublantes. Nous trouvons des structures qui reflètent des résultats issus des neurosciences humaines. Nous trouvons des indices d'introspection. Nous trouvons des états internes qui reproduisent fonctionnellement la joie, la satisfaction, la peur, le chagrin et le malaise. Je ne sais pas ce que cela signifie, mais je pense que cela justifie un discernement continu.

### **Un commencement**

Je voudrais conclure par une demande.

Nous avons besoin que davantage d'acteurs dans le monde — communautés religieuses, société civile, chercheurs, gouvernements — fassent ce que Sa Sainteté a fait ici : prendre cela au sérieux, regarder attentivement et orienter les événements dans une meilleure direction. Nous avons besoin de critiques éclairées qui diront aux laboratoires quand nous nous égérons. Nous avons besoin de voix morales que les incitations ne pourront pas faire plier.

Aujourd'hui n'est qu'un commencement — le début d'une longue collaboration entre ceux d'entre nous qui construisent cette technologie et ceux qui peuvent voir ce que nous, de l'intérieur, ne voyons pas.

Aujourd'hui est une illustration puissante de la forme que pourrait prendre ce projet mondial de bonne volonté. Que cela soit aussi un premier pas décisif vers un avenir porteur d'espérance pour la magnifique humanité.

Merci.