

Avis sur « The Case for Boycotting Generative AI » de Gary Marcus

Réponse à une demande d'avis de groupe

1. Marcus lui-même : un point sur sa crédibilité

Déjà Marcus lui-même a (jusqu'à genre, exactement maintenant) depuis longtemps minimisé les capacités des systèmes d'IA, notamment en bornant sa critique aux LLMs seuls sans prendre en compte leur environnement d'utilisation (ce à quoi ils sont reliés, les outils auxquels ils ont accès).

Donc à la fois, cool son avis évolue, à la fois c'est pas le gars auquel je fais le plus confiance sur le sujet. Il a quand même le bon goût de pas être en lien d'intérêts avec les gros labos d'IA (en tout cas il n'y bosse pas, n'a pas comme plan de carrière d'y bosser, et les critique), ça reste un bon point pour lui parce que ça reste un très petit monde.

2. Sur le constat : d'accord, mais la distinction n'est pas opérationnellement utile

Sur le constat je suis globalement d'accord, le risque existe, il est aujourd'hui principalement lié aux agents (LLM+harness) plutôt qu'aux LLM en soit, il est amplifié fortement par leur accès à internet. Je ne pense cependant pas que ces distinctions soient opérationnellement utiles : si un LLM a accès à un terminal, on peut considérer qu'il a accès à internet (vu les capacités en termes de cybersécurité). On pourrait penser à l'isolation physique, mais aujourd'hui dans toute organisation de plus de 10 personnes c'est vachement dur à garantir. Le truc peut assez facilement convaincre de le laisser avoir accès à internet : « si j'avais accès à internet, regarde tout ce qu'on pourrait faire! » et a chair est faible.

BREF même si les risques viennent bien de lier des LLM à des terminaux et à leur donner accès à internet, en pratique ce qui pose souci c'est tout simplement que les LLM soient capables d'utiliser un terminal, et ils le sont.

3. Auto-régulation des labos et accord US–Chine : espoir quasi nul

Je partage aussi un espoir quasi nul que les labos se modèrent tout seuls, et un espoir tout aussi faible pour un accord en bonne foi ou réellement coercitif entre US et Chine. Même en pratique je vois pas ce que ça ferait ce genre d'accord, à la limite il y aurait peut être moyen d'imposer aux fabricants hardware une limitation à ce niveau, ça empêcherait de faire tourner des modèles d'IA sur du nouveau matos ou d'en entraîner de plus puissants, en tout cas ça ferait une forte réduction mais bon, même techniquement ça a pas l'air simple.

4. Sur sa « solution » : le boycott

Sur sa « solution » boycott ça me paraît inutile pour plein de raisons. En supposant qu'un tel boycott soit possible avec une envergure suffisante (ie que les entreprises utilisatrices se joignent à l'effort lol) il y a en effet moyen de limiter la capacité des gros labos à entraîner des modèles de plus en plus puissants, mais le truc c'est que des modèles suffisamment puissants pour être dangereux sont déjà dispos en poids ouverts.

Après voilà j'ai des a priori négatifs sur les boycotts, en particulier lorsqu'ils concernent des produits vendus aux entreprises. Pour moi le boycott c'est plus un truc qu'on fait pour soi, pour ses valeurs personnelles, qu'une solution politique. Je dois dire que dans ce cas là ça me rappelle aussi un peu le côté responsabilisation individuelle qu'on voit vis à vis du changement climatique.

5. Alors les solutions, c'est compliqué

En dehors de :

- destruction des capacités de production énergétiques
- abattre le capitalisme

Côté régulation, je pense que le truc le plus sensé c'est de clarifier la responsabilité légale (et financière) pour tous les problèmes qui touchent à l'ia. Genre, quand une IA fait une connerie, qui est responsable légalement? Ça fonctionnerait très bien de dire que Sam Altman (les patrons quoi) ou bien ceux qui détiennent du capital de la boîte, sont pénalement responsables des agissements de leurs ia (dans le cas de huggingface ça ferait de trèèèèèèèè nombreuses charges de piratage assez graves).

Bon j'y crois pas trop donc le truc plus réaliste ça serait malheureusement de taper sur l'utilisateur final. Le souci c'est qu'il n'y en a aps toujours (ya de plus en plus d'IA qui paient leurs propres tokens de fonctionnement et qui agissent en autonomie). Breeeeeef ya plein d'autres trucs mais j'en ai marre d'écrire.

L'avis de Jérémy Perret sur le sujet sera bieeeeeennnnnnnn plus intéressant que le mien donc j'arrête de vous polluer.

Bonne journée !