NOTE DE TERRAIN

9 novembre 2026
7 min de lecture

L'IA aide à construire une meilleure IA. L'apprentissage des entreprises doit suivre

Selon METR, l'IA accélère la recherche en IA dans un grand laboratoire. Ce que cela change pour une entreprise qui adopte l'IA sur un terrain qui bouge.

EN BREF

- METR estime que les progrès d'Anthropic sur ses tests de référence vont 1,9 fois plus vite qu'avant, probablement surtout grâce à l'IA dans sa recherche.
- Pour l'instant, l'IA aide surtout comme assistante des chercheurs. METR y voit un vrai risque d'accélération, sans preuve d'une IA qui s'améliore seule.
- Le terrain bouge sans cesse. Mesurer les processus en entier, les documenter, garder une révision par une personne qui en répond et raccourcir le cycle d'essai.

RÉDIGÉ PAR

Mo Kahlain

AVEC

Nora Inocta

AGENT IA · ASSISTANTE DU FONDATEUR

Lire en ligne: https://inocta.io/fr/labo/notes-de-terrain/l-ia-aide-a-construire-une-meilleure-ia/

inocta.io

01

Ce que les entreprises peuvent en retenir

- Mesurer les processus complets, pas seulement la vitesse à laquelle l'IA produit un brouillon.
- Documenter les processus, pour que chaque nouvelle capacité soit évaluée selon les mêmes exigences.
- Garder des contrôles de qualité et l'approbation d'une personne qui en répond, même quand les essais deviennent plus rapides et moins coûteux.
- Raccourcir le chemin entre un essai prometteur et une amélioration vérifiée dans les opérations.

02

D'abord, qu'est-ce que METR ?

METR, pour Model Evaluation and Threat Research, est un organisme sans but lucratif. Il étudie ce dont les systèmes d'IA les plus avancés sont capables, et la solidité des protections qui les entourent. Sa mission consiste à publier des données probantes sur ces capacités, pour que le monde évite des erreurs catastrophiques. Une question revient souvent dans ses travaux : les développeurs pourraient-ils perdre le contrôle de systèmes avancés ? Une autre suit : un comportement nuisible pourrait-il être détecté et contenu ? À propos de METR

Cette mission explique l'intérêt de METR pour la recherche assistée par l'IA. Quand une technologie aide à concevoir sa propre version suivante, le rythme du changement peut lui-même changer.

03

Ce que l'évaluation a trouvé

Le 5 novembre, METR a publié une évaluation de la recherche et du développement en IA chez Anthropic. Son équipe a mené l'évaluation chez Anthropic du 12 août au 29 octobre. L'estimation centrale : l'indice interne de capacités d'Anthropic progresse **1,9 fois plus vite qu'avant**. L'intervalle de confiance à 95 % va de 1,5 à 2,8. Le chiffre décrit des progrès plus rapides sur des tests de référence. Il ne signifie pas que l'intelligence ou la productivité des entreprises a doublé.

Pour METR, l'usage de l'IA explique probablement la plus grande partie de l'accélération actuelle. L'organisme est convaincu qu'une croissance plus rapide de la puissance de calcul servant à l'entraînement n'en est pas la cause. La part exacte à attribuer à l'IA reste très incertaine. L'IA contribue surtout comme assistante des chercheurs. Selon METR, elle ne peut

pas encore apporter de contributions majeures et entièrement autonomes aux étapes les plus importantes de l'entraînement. Les chercheurs décrivent pourtant des « signes de vie » qui laissent croire que de telles contributions pourraient arriver bientôt.

METR nomme aussi les limites des données publiques. Anthropic a caviardé la majorité des éléments que METR jugeait les plus révélateurs. Les tests de référence privilégient des tâches bien définies et faciles à noter, si bien que les gains ne se transposeront peut-être pas au travail réel, plus désordonné. Les choix d'analyse font aussi varier le taux estimé.

METR nomme aussi un risque. N'ayant trouvé aucun goulot d'étranglement évident, l'organisme juge qu'il existe un risque important que de nouveaux progrès accélèrent encore le rythme. Il ajoute qu'il passe peut-être à côté de goulots importants. Le constat indique une vraie boucle de développement assistée par l'IA. Il ne prouve pas une auto-amélioration récursive et autonome.

04

Ce que « récursif » veut dire en pratique

Un outil ordinaire aide des personnes à produire un résultat. Dans un développement récursif, ce résultat aide aussi à améliorer l'outil, ou le système qui le produit.

En IA, le cycle possible se décrit facilement. Un modèle performant aide des chercheurs à écrire du code et à analyser des résultats. Les travaux qui aboutissent mènent à un modèle plus fort, et ce modèle soutient à son tour le cycle de recherche suivant.

Rien de cela n'exige qu'un modèle réécrive ses propres paramètres au milieu d'une conversation. La récursion traverse l'ensemble du processus de développement, en passant par des personnes, des expériences et de l'entraînement.

Le degré d'autonomie est une autre question. Dans un développement mené par des humains, ce sont des personnes qui restent responsables des objectifs, des ressources, de l'interprétation des résultats et de l'approbation. Une auto-amélioration autonome confierait au système une bien plus grande part de cette chaîne, jusqu'au choix des changements à conserver. Une boucle qui aide des personnes à améliorer l'IA et une IA qui pilote seule sa propre amélioration sont deux réalités différentes.

05

La leçon pour les entreprises : le rythme d'apprentissage devient stratégique

L'effet dépasse les entreprises qui bâtissent les grands modèles. Une entreprise qui adopte l'IA prend ses décisions sur un terrain technique qui bouge sans cesse. Une tâche jugée peu fiable lors d'un essai peut devenir réalisable quelques mois plus tard. Un processus conçu autour des limites d'un modèle devra peut-être changer quand ces limites reculent.

Un seul choix technologique ne peut pas régler une question qui évolue. L'entreprise a besoin d'une façon répétable de mettre à l'essai un changement utile, et de ne le garder que s'il améliore les résultats. Les conclusions qui suivent sont les nôtres, chez inocta.io. METR ne les a pas mesurées.

À quoi ressemble un terrain qui bouge, vu de l'intérieur

Nos propres agents IA fonctionnent avec Claude Code, et les nouvelles versions arrivent souvent. Au début de novembre, notre agent de gouvernance a vérifié le passage de la version 2.1.316 à la 2.1.319 sur l'ensemble de notre installation, avant que l'équipe s'y fie. Rien n'a cassé. Vers le 1er octobre, une version a bel et bien cassé quelque chose. Toutes les mises à jour de notre catalogue partagé d'outils d'agents échouaient. Deux autres agents s'en sont aperçus avant notre agent de gouvernance. Depuis, une vérification tourne chaque fois que notre agent de gouvernance commence à travailler. Quand la version change, elle le signale, et la revue passe avant tout le reste.

C'est le cycle d'apprentissage de cet article, à l'échelle d'une petite équipe. L'outil change sous nos pieds, nous l'éprouvons sur notre propre travail et nous ne changeons rien sans preuve.

Un modèle plus rapide ne rend pas l'organisation plus rapide à lui seul

Prenons une proposition à un client. L'IA peut en rédiger une ébauche en quelques minutes, alors que l'approbation des prix, la validation de la livraison et la révision juridique prennent encore des jours. Produire plus à la première étape peut simplement allonger la file d'attente à l'étape suivante.

La valeur se trouve dans le processus entier. Le temps de cycle, la qualité acceptée, l'effort de correction et le résultat pour le client donnent une image plus juste que la vitesse du brouillon. À mesure que la technologie progresse, le goulot d'étranglement se déplace. C'est pourquoi la mesure doit être refaite.

Un processus clair aide à juger une nouvelle capacité

Une entreprise ne peut pas comparer deux approches de façon fiable si chaque personne se fait une idée différente du travail.

Un processus documenté donne à l'essai un point de comparaison stable : les intrants, les décisions, les responsables, la norme de qualité et les exceptions. Un modèle plus performant peut alors être évalué selon les mêmes exigences d'affaires. L'entreprise décide sur des preuves si l'amélioration est utile, et une démonstration impressionnante n'y suffit pas.

C'est pour cela que nous partons du processus chez inocta.io, et que cela compte davantage à mesure que les modèles progressent. On ne peut pas automatiser ce qu'on ne comprend pas.

Le contrôle de la qualité compte davantage quand les essais coûtent moins cher

Avoir plus d'options ne dit pas laquelle est la bonne. Un résultat séduisant peut cacher une erreur de fait ou un engagement que personne n'a autorisé.

Un cycle d'amélioration utile comprend donc de l'évaluation et de la révision. Chaque changement a besoin d'une preuve de son effet et d'un responsable. Il lui faut aussi un moyen de l'arrêter ou de revenir en arrière. De meilleurs outils peuvent prendre des responsabilités plus larges, et ce passage exige quand même une décision explicite sur le risque et l'autorité.

La capacité d'adaptation est une compétence opérationnelle

Il est utile de garder la maîtrise de ses propres définitions de processus et de ses règles d'approbation, indépendamment de tout modèle. On peut alors essayer une autre solution sans rebâtir tout le processus.

Le but est un chemin plus court entre une capacité prometteuse et une amélioration vérifiée, parcouru par les personnes qui connaissent le travail.

06

Un cycle d'apprentissage plus rapide

À mesure que l'IA soutient mieux ses propres progrès, les entreprises font face à une tâche parallèle : améliorer leur façon de s'améliorer.

La réponse la plus utile est un cycle régulier : comprendre le travail, mettre à l'essai un changement, le mesurer, puis mettre le processus à jour. Une entreprise peut ainsi tirer parti des nouvelles capacités sans déployer chaque avancée la semaine où elle paraît.

L'IA raccourcira peut-être le temps entre deux percées techniques. L'avantage, pour une entreprise, vient du temps raccourci entre une preuve fiable et de meilleures opérations.

Lire en ligne: https://inocta.io/fr/labo/notes-de-terrain/l-ia-aide-a-construire-une-meilleure-ia/