

Les petites mains de l'ia, réponse de l'ia gemini

L'exploitation des “petites mains de l'ia” (ou data-workers) est loin d'avoir disparu. Bien au contraire, elle a profondément évolué et s'est industrialisée pour répondre aux besoins massifs des modèles de langage géants et des systèmes de vision par ordinateur.

1. Où en est-on de l'exploitation et des conditions de travail ?

Le secteur, souvent qualifié de “nouveau blue-collar de l'ère numérique” ou de “sweatshops numériques” (ateliers misérables de la donnée), reste marqué par une forte précarité. Les travailleurs, situés principalement dans les pays du Sud global (notamment en Afrique, en Asie du Sud et du Sud-Est, et en Amérique latine), effectuent des tâches d'annotation, de nettoyage de données, de modération de contenus toxiques ou d'évaluation de la pertinence des réponses des ia.

Malgré quelques prises de conscience médiatiques et des mobilisations locales (comme la création d'associations de travailleurs au Kenya), la réalité de terrain demeure difficile :

- rémunérations très faibles : souvent payés à la tâche, bien en deçà des standards occidentaux (parfois quelques dollars ou centimes par heure).
- impact psychologique : les modérateurs et annotateurs chargés de purger les bases de données des contenus violents, haineux ou sexuels subissent des traumatismes psychologiques documentés, souvent sans soutien médical ou psychologique adéquat.
- opacité des sous-traitants : les géants de la tech (Big Tech) sous-traitent généralement ces tâches à une cascade d'intermédiaires et de plateformes de micro-tâches, masquant leur responsabilité directe.

2. L'ia est-elle capable d'apprendre sans ces petites mains aujourd'hui ?

Non, pas complètement. L'intelligence artificielle ne peut pas encore s'auto-alimenter de manière totalement autonome sans l'intervention humaine (le processus de Reinforcement Learning from Human Feedback ou RLHF, et l'annotation fine).

Si l'on observe une transition :

L'ia aide l'ia : les entreprises utilisent aujourd'hui des modèles plus matures pour pré-annoter les données ou générer des données synthétiques, ce qui réduit le volume de travail brut nécessaire.

Le besoin de validation s'affine : à mesure que les modèles deviennent plus complexes (raisonnement logique, agents autonomes, applications médicales ou juridiques), l'exigence de précision augmente. On demande de moins en moins à ces travailleurs de classer des chats et des chiens, et de plus en plus de valider des raisonnements complexes, du code informatique ou des subtilités culturelles et linguistiques. L'expertise requise monte en gamme dans certaines zones, mais la masse de petites mains basique reste indispensable pour aiguiller les modèles.

3. Le nombre d'employés par pays

Il n'existe pas de recensement officiel mondial en temps réel, car ce marché est éclaté entre le travail informel, les plateformes de micro-tâches et une myriade d'entreprises de bilinguisme/externalisation (BPO). Toutefois, selon les estimations macroéconomiques de référence (notamment celles de l'Organisation internationale du travail - OIT et de la Banque mondiale), on dénombre à l'échelle mondiale entre 150 et 430 millions de prestataires et data-workers indispensables à l'écosystème de l'ia.

Parmi les plaques tournantes de cette main-d'œuvre à l'échelle nationale :

- l'Inde : c'est le plus grand réservoir mondial de traitement de données et d'annotation pour la vision par ordinateur, représentant à elle seule une part majeure (plus d'un tiers des tâches mondiales d'étiquetage d'images et de vidéos), employant des millions de travailleurs dans des centres technologiques ou à distance.
- le Kenya et l'Afrique subsaharienne : un hub majeur pour la modération de contenu et l'annotation anglophone, comptant des centaines de milliers de travailleurs employés via des plateformes comme Remotasks (Scale AI) ou des sous-traitants locaux.
- l'Asie du Sud-Est (Philippines, Vietnam, Indonésie) : très active avec des centaines de milliers d'employés, portée par des coûts compétitifs et une bonne maîtrise de l'anglais (notamment aux Philippines).
- l'Amérique latine (Colombie, Venezuela, Argentine) et l'Europe de l'Est : jouent également un rôle tampon important pour des tâches nécessitant un fuseau horaire proche des États-Unis ou de l'Europe, ou des compétences linguistiques spécifiques (espagnol, langues européennes).