

Préserver l'autonomie de pensée et la maîtrise des données dans un environnement assisté

L'Intelligence Artificielle progresse vite, souvent plus vite que notre capacité collective à en mesurer les implications. Derrière les prouesses techniques, un enjeu plus discret mais déterminant prend forme : l'avenir de notre manière de raisonner, décider et protéger ce qui fait la valeur des entreprises : leurs données et leur intelligence collective.

Cette évolution soulève aussi la question de l'intégrité cognitive, autrement dit, notre capacité à conserver une pensée libre, non manipulée, qui ne se laisse pas happer par les « faits alternatifs », les deepfakes ou les distorsions discrètes que peuvent générer ou amplifier certains systèmes automatisés. Et ce n'est pas qu'un débat théorique, en Suisse, [une initiative populaire](#) déposée au Parlement fédéral propose de mieux encadrer cette intégrité cognitive. Preuve que la frontière entre information, opinion et production algorithmique devient suffisamment poreuse pour nécessiter un vrai débat de société.

À cela s'ajoutent les avancées d'acteurs comme « Apertus », premier grand modèle de langage open source suisse. Malgré quelques critiques sur ses performances, les chercheurs rappellent que sa vocation n'est pas de rivaliser avec les géants propriétaires, mais d'offrir une base ouverte, auditable et responsable, pensée pour servir l'intérêt général.

On observe que, dans beaucoup d'entreprises, l'IA est utilisée de manière très opportuniste et dans une logique de rendement pour rester compétitif : on teste, on explore, on copie-colle des idées, on envoie des documents « pour voir ». Derrière des gains rapides... s'ouvre aussi des zones de fragilité souvent sous-estimées, notamment en matière de protection des données.

À ce stade, beaucoup d'utilisateurs·rices alimentent des IA grand public avec des informations sensibles sans mesurer où elles vont, dans quelles bases elles sont stockées, ni comment elles pourraient être réutilisées pour entraîner les modèles.

À cette vulnérabilité s'ajoute une autre dérive : la non-remise en question. L'IA répond vite, avec aplomb, et beaucoup s'arrêtent là. Le problème ne tient pas seulement aux erreurs factuelles fréquentes, mais à la manière même dont ces systèmes produisent leurs réponses. Ils ne cherchent pas la vérité : ils calculent des probabilités de cohérence. Or, les sources mobilisées restent souvent opaques, agrégées, puis transformées par différentes couches invisibles : filtres, optimisations, techniques génératives, qui modifient l'information sans que l'utilisateur ne perçoive les biais introduits. Ces biais, parfois discriminatoires, sont déjà bien documentés dans la recherche et les médias (notamment par [l'Université de Genève](#) ou encore dans une étude de l'Université de Zurich relayée par [l'Agefi](#)).

Enfin, un point rarement évoqué mérite d'être dit : le contenu généré par l'IA porte aussi l'empreinte de ceux qui la conçoivent, la financent et la gouvernent. Les systèmes ne sont jamais totalement neutres : ils reflètent, volontairement ou non, les intentions, les angles morts et les priorités de leurs concepteurs.

Pour les entreprises, la priorité consiste donc à maintenir une capacité de jugement critique, même face à des analyses très convaincantes, et à protéger les données internes, les éléments confidentiels, ainsi que tout ce qui constitue l'avantage compétitif. Autrement dit : garder la main pour que la technologie génère réellement de la valeur.

Concrètement, cela passe par quelques réflexes simples mais déterminants : sensibiliser les équipes à la protection des données avant tout usage d'outils externes ; instaurer un cadre clair incluant la validation humaine, la vérification systématique des faits et l'interdiction d'introduire des informations sensibles dans des systèmes non maîtrisés ; renforcer la capacité des collaborateurs·rices à vérifier, comparer et questionner les réponses générées par l'IA ; identifier

précisément les zones à risque (recrutement, évaluation, finance, communication), où l'usage doit être particulièrement encadré ; et enfin, préserver des moments de réflexion non assistée pour éviter qu'une dépendance cognitive ne s'installe sans qu'on s'en rende compte...

PS : pour la transparence, ce texte est le fruit d'une co-crédation entre l'IA et ma pensée : les sources ont été injectées dans le modèle, puis l'ensemble a été relu, corrigé et challengé côté humain, parce que l'assistance aide, mais la pensée humaine reste irremplaçable.

Adélaïde Gonzalez
emmenegger compétences conseils sa