INSTITUT Montaigne

institutmontaigne.org
NAC2401-01

**NOTE D'ACTION** - Janvier 2024

# Pour une Autorité française de l'IA


## L'urgence d'un espace de compétences et de recours sur l'IA

La France accueillera bientôt la deuxième grande édition du Sommet Mondial de l'IA, initié par le Royaume-Uni en novembre 2023. Néanmoins, notre pays manque d'un outil de gouvernance indispensable sur son propre territoire et comme maillon européen : une Autorité française de l'IA. Cette Autorité aurait pour mission principale d'évaluer rigoureusement la performance et les risques de l'état de l'art de la technologie, qui avance à grands pas, et d'accompagner les entreprises de l'IA dans leur gestion des risques. Depuis la publication de la note d'action de l'Institut Montaigne d'avril 2023 sur l'IA « sûre et digne de confiance » qui formulait des recommandations en ce sens, plusieurs pays *leaders* de l'IA – le Royaume-Uni, les États-Unis et Singapour – ont annoncé leur propres Instituts de Sûreté de l'IA en fin d'année 2023. La France est désormais en retard. Une première étape est de créer, dès 2024, un Centre d'évaluation de l'IA, rassemblant les forces et acteurs en présence. Un tel centre pourra préfigurer une future Autorité créée par la loi, à même de faire avancer les sujets d'IA en France en plus de devenir un interlocuteur facilitant la mise en conformité de l'application du règlement européen en la matière.

## L'évaluation : un outil clé pour bâtir la gouvernance mondiale émergente

Lors du premier Sommet Mondial de l'IA en novembre 2023, 28 pays, dont la France, se sont engagés à maîtriser l'IA *via* deux leviers : l'identification et l'évaluation des risques, d'une part, et des mesures de gouvernance centrées sur les risques, d'autre part.

Le Royaume-Uni et les États-Unis ont rapidement adopté une approche centrée sur l'évaluation des risques pour la sécurité nationale des États de l'IA à l'état de l'art (*Frontier AI* en anglais) – soit les quelques grands modèles d'IA à usage général dépassant un seuil de performance particulièrement élevé – sans toutefois imposer de mesures de gouvernance *a priori*.

D'autres acteurs se sont dotés de règles contraignantes, en estimant que l'auto-régulation par les entreprises n'était pas satisfaisante. C'est le cas de l'Europe, qui régulera avec son *AI Act* les systèmes d'IA déployés dans des cas d'usages à « risque élevé » et les plus grands modèles d'IA à usage général, indépendamment de leur cas d'usage. C'est aussi le cas de la Chine, qui a rapidement adopté une série de mesures pour strictement réguler l'IA et les acteurs qui commercialisent des systèmes d'IA à usage général. En octobre 2023, Pékin a également introduit un cadre de gestion des risques pour la recherche dans l'IA, lui permettant de distinguer plus facilement les enjeux de contrôle du contenu généré par l'IA et les risques liés à la recherche sur l'IA de pointe.

Qu'il existe un cadre réglementaire ou non, l'évaluation apparaît désormais comme un outil indispensable : il peut être un filet de sécurité minimum en l'absence de règles strictes, permettant une intervention spontanée des pouvoirs publics en cas de risque jugé inacceptable ; il peut être également le moyen de faire évoluer la réglementation au rythme de la technologie.

## Anticiper la déferlante IA et les risques associés

L'année 2023 aura été celle de la déferlante IA, marquant une accélération sans précédent de l'évolution de la technologie. Avec ChatGPT, le grand public a découvert la puissance d'une nouvelle ère de l'IA, celle des systèmes d'IA à usage général, capables d'effectuer un grand nombre de tâches différentes (coder, dialoguer en plusieurs langues, résoudre des problèmes de physique ou de chimie, etc.).

Depuis, la technologie a fait des bonds significatifs. En mars 2023, OpenAI, l'entreprise mère de ChatGPT, a dévoilé son modèle GPT-4, dont le Quotient intellectuel (QI) mesuré par des tests d'aptitude intellectuelle dépasse celui d'entre 80 % et 99 % de la population. Les capacités d'action de ces modèles d'IA ont rapidement été augmentées au-delà de la simple génération de textes ou d'images. En particulier, ils ont été utilisés pour créer des agents IA, c'est-à-dire des systèmes d'IA dotés d'objectifs, connectés à d'autres outils externes tels que des outils de codage informatique, des moteurs de recherche, des laboratoires scientifiques pilotés à distance *via* internet et donc désormais capables d'interagir avec le monde externe en toute autonomie.

Chacun commence à comprendre le potentiel transformatif de cette technologie, constate le rythme exponentiel de son développement et ressent le besoin de mieux l'appréhender. L'enjeu est désormais de passer à l'action.