Adoption de l'AI Act : promesses et ambitions de la première législation occidentale sur l'intelligence artificielle

Caroline Lequesne Roth

Adoption de l'AI Act : promesses et ambitions de la première législation occidentale sur l'intelligence artificielle

Caroline LEQUESNE, maître de conférences HDR en droit public, Université Côte d'Azur

Ceci est la version longue (et antérieure) d'un entretien paru au *Recueil Dalloz*, D. 2024, p. 864.

Le printemps 2024 entérine trois années de discussions intenses, qui ont vu naître la première réglementation sur l'IA du monde occidental. Après l'accord politique obtenu le 8 décembre 2023, le vote du Parlement européen le 13 mars dernier, c'est au Conseil qu'appartiendra le vote final en mai prochain. Face aux enjeux et aux intérêts pluriels qui ont animé le débat législatif quelle approche a permis de sceller le consensus européen ?

Entre les espoirs d'un « *Brussel effect* », les craintes pour les libertés fondamentales et le souci de préserver la compétitivité européenne, ce texte noue des compromis composites, suscitant insatisfactions et perplexités des camps adverses. Cela tient à sa nature hybride, que Juliette Sénéchal décrit comme « l'hydre à trois têtes » : le texte répond à des injonctions nombreuses, souvent contradictoires, de la sécurité des produits aux enjeux « civilisationnels » que l'arrivée des IA génératives a contribué à mettre en lumière. Le législateur européen n'en a pas moins retenu une approche commune, fils rouge du texte et du système qu'il battit : l'approche par les risques. Celle-ci confère un point de départ ; ce n'est pas tant la technologie elle-même que ses potentiels effets qui constituent le fait générateur des obligations qui s'imposent aux déployeurs et fournisseurs d'IA (ci-après, les « opérateurs »). Les risques se déclinent ainsi en catégories, des IA aux risques inacceptables - et par suite interdites - aux IA à haut risque et autres risques « systémiques » qui font l'objet d'un encadrement renforcé. La première catégorie, inscrite à l'article 5, regroupe les systèmes d'IA susceptibles de manipuler, exploiter les vulnérabilités d'un individu ou le catégoriser sur la base d'un comportement social, d'un score inféré de celui-ci ou de ses émotions ; la catégorie intègre également les systèmes d'identification biométrique à distance, « en temps réel », dans des espaces accessibles au public que les exemptions laissent toutefois largement exploitables. En dépit des débats suscités par ces premiers systèmes, dans leur usage par les forces de l'ordre notamment, l'effort réglementaire se concentre sur la seconde catégorie. Les opérateurs de systèmes d'IA à haut risque, identifiés notamment dans les secteurs considérés comme sensibles (éducation, emploi, infrastructure sensible, service public, forces de l'ordre, contrôles aux frontières, administration de la justice) seront soumis une série d'obligations générales et sectorielles (obligations de transparence, management des risques, recherche de biais en matière de matière de gouvernance des données, obligation de préserver un contrôle humain, analyse d'impact). Le non-respect de

celles-ci est passible de sanctions, qui pourront s'élever jusqu'à 35 millions d'euros ou 7 % du chiffre d'affaires consolidé de l'entreprise concernée pour les violations des applications d'IA interdites, à 15 millions d'euros ou à 3 % pour les violations des obligations entourant les systèmes d'IA à haut risque et à 7,5 millions d'euros ou à 1,5 % pour la communication d'informations inexactes.

Comment le texte répond-il aux enjeux particuliers de l'IA générative ?

La diffusion à échelle d'IA génératives fondées sur de larges modèles de langage - à l'instar de Chat GPT, Midjourney ou HeyGen - a marqué un tournant dans les négociations européennes. Si un « chat » ne constituait pas, en soi, un système hautement risqué conformément à la grille analytique proposée, il n'en menaçait pas moins nos fonctionnements démocratiques (désinformation, atteinte au droit d'auteur, production de faux contenus intimes, « *deepfakes* », etc.). Pour y répondre, le texte a intégré, sous l'impulsion du Parlement européen, un régime spécifique aux modèles d'IA à usage général. Ces modèles, entraînés à l'aide d'une grande quantité de données en utilisant l'autosupervision à grande échelle, ont des « capacités à fort impact » et présentent un « risque systémique au niveau de l'Union », ayant « une incidence significative sur le marché intérieur en raison de sa portée, et des effets négatifs réels ou raisonnablement prévisibles sur la santé publique, la sécurité publique, les droits fondamentaux ou la société dans son ensemble, qui peut se propager à grande échelle dans toute la chaîne de valeur » (article 4). Les obligations mises à la charge des fournisseurs des modèles d'IA à usage général sont fortement inspirées du *Digital Service Act* (DSA) (Règl. [UE] 2022/2065 du 19 oct. 2022 relatif à un marché unique des services numériques et modifiant la dir. 2000/31/CE, JO L 277 du 27 oct. 2022), qui impose notamment aux très grands moteurs de recherche et plates-formes des obligations supplémentaires pour mitiger les risques systémiques liés à leurs systèmes de modération et de recommandation. L'AI Act prévoit, de même, des obligations de transparence spécifiques à la charge des fournisseurs de modèles d'IA à usage général (tenue à jour de la documentation technique ; mise à disposition du public d'un résumé présentant les bases d'apprentissage du modèle) et du respect du droit de l'Union européenne en matière de droit d'auteur (article 53).

Quel modèle de gouvernance doit assurer la mise en œuvre du texte ?

Dans la veine des grands textes européens du numérique, la mise en œuvre de l'AI Act, incrémentale, reposera sur un système de corégulation complexe à deux échelons. Au niveau national, il appartiendra à chaque Etat membre de désigner les autorités compétentes : une autorité de notification des organismes évaluateurs de la conformité, et « au moins » une autorité de surveillance du marché, chargée d'effectuer des inspections, pour les IA à haut risque. Le pragmatisme requis face à la surenchère institutionnelle justifie que des autorités existantes, à l'instar de la CNIL, répondent à ces fonctions. A l'échelon européen, harmonisation, coordination et mise en œuvre prendront appui sur deux acteurs majeurs : le Bureau

européen de l'IA, créé au sein de la Commission, qui assurera notamment l'interprétation des dispositions relatives aux modèles d'IA à usage général, et le Comité européen de l'intelligence artificielle composé d'un représentant par État membre, auquel se joindront, sans droit de vote, le contrôleur européen de la protection des données et les organismes ou experts nationaux et de l'Union européenne, en fonction des problématiques traitées. Un groupe scientifique d'experts indépendants accompagnera ses instances dans la mise en œuvre du règlement (alerte sur les risques systémiques, élaboration d'outils et de méthodes d'évaluation, conseils sur les classifications). Le texte requiert enfin la proactivité des opérateurs, garants de leur mise en conformité (auto-évaluation/auto-régulation) par le respect, notamment, des normes techniques et des procédures de certification.
L'architecture ambitieuse de ce modèle en constitue aussi la potentielle faiblesse : les moyens financiers et humains alloués aux autorités, la traduction des exigences fondamentales en normes, le jeu des entreprises sont autant d'inconnues qui détermineront l'efficacité de cette première législation.