

PROCESSUS DU G7, DIT D'HIROSHIMA, SUR L'INTELLIGENCE ARTIFICIELLE (IA) GÉNÉRATIVE

VERS UNE VISION COMMUNE À L'ÉCHELLE DU G7 SUR L'IA GÉNÉRATIVE

RAPPORT ÉTABLI
SOUS LA PRÉSIDENTE
JAPONAISE DU G7
ET DU GROUPE DE
TRAVAIL DU G7 SUR
LE NUMÉRIQUE ET LA
TECHNOLOGIE

7 septembre 2023

Les dirigeants du G7 ont recensé les thèmes à examiner dans le cadre du processus d'Hiroshima et demandé l'établissement rapide d'un état des lieux des possibilités et enjeux associés à l'IA générative. Le présent rapport donne à voir les réponses reçues à un questionnaire conçu pour faciliter cet état des lieux et doit étayer les échanges du G7 au sujet des priorités d'action communes en ce qui concerne cette technologie. Il permet également de brosser une rapide vue d'ensemble de son évolution dans le temps et dans l'espace. Le rapport et les réponses au questionnaire doivent être considérés comme le reflet de la situation à un instant T : ils donnent une idée des tendances observables à l'été 2023, dans un contexte technologique en mutation rapide. Le rapport a contribué à éclairer et structurer les discussions tenues dans le cadre du processus du G7, dit d'Hiroshima, sur l'IA.

Le présent document a été établi par la Direction de la science, de la technologie et de l'innovation (STI) de l'Organisation de coopération et de développement économiques (OCDE) pour le compte de la Présidence japonaise du G7, au titre de l'année 2023, et celui du Groupe de travail du G7 sur le numérique et la technologie, en vue d'éclairer les échanges tenus dans le cadre du Processus du G7, dit d'Hiroshima, sur l'IA et dans celui de la Réunion virtuelle intermédiaire des ministres sur l'IA, le 7 septembre 2023. Les opinions qui y sont exprimées et les arguments qui y sont avancés ne reflètent pas nécessairement les points de vue officiels des pays membres de l'OCDE ou du G7.

Le présent document est publié sous la responsabilité du Secrétaire général de l'OCDE.

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Image de couverture : ©TippaPatt/Shutterstock.

© OCDE 2023

L'utilisation de ce document, sous forme numérique ou imprimée, est régie par les conditions générales d'utilisation consultables à l'adresse : <http://www.oecd.org/fr/conditionsdutilisation/>.

RÉSUMÉ

L'IA générative s'est rapidement imposée sur la scène publique.

Depuis que le public a eu vent de son existence, l'IA générative est chaque jour plus présente dans la conversation de tout un chacun, partout dans le monde. Le nombre d'articles de journaux et de tweets qui lui est consacré a été multiplié par huit en l'espace de six mois à peine.

Les progrès de la recherche dans le domaine de l'IA générative, notamment ceux du développement de code source libre, ont été précédés d'une forte augmentation des investissements.

La vaste publicité faite à l'IA générative et son adoption rapide ont été permises par les progrès constants de la recherche et du codage. Des innovations fondamentales, comme les modèles de type *Transformers* (appelés aussi modèles transformateurs ou d'auto-attention), conjugués à l'augmentation de la puissance de calcul, ont ouvert la voie à la prolifération des grands modèles de langage ainsi qu'à d'autres types de modèles d'IA générative. Les publications scientifiques et le développement de code source libre d'IA générative ont, en parallèle, connu un essor remarquable depuis 2017, qui s'est encore accentué en 2023. Les investissements en capital-risque se sont envolés : leur montant était estimé à 12 milliards USD, à l'échelle mondiale, rien que pour le premier semestre 2023. Les publications scientifiques et le développement de code source libre en relation avec l'IA générative ont, en parallèle, connu une croissance prodigieuse depuis 2017, qui s'est encore accentuée en 2023.

Les progrès rapides de l'IA générative s'expliquent par le potentiel attendu de celle-ci, entre gains de productivité et soutien à l'innovation et à l'entrepreneuriat, sans oublier l'apport de solutions aux enjeux mondiaux.

En réponse à un questionnaire d'enquête qui leur a été communiqué au T3 2023, les membres du G7 ont, à l'unanimité, fait savoir qu'ils considéraient les gains de productivité, la promotion de l'innovation et de l'entrepreneuriat et la concrétisation de solutions aux enjeux mondiaux comme quelques-unes des principales possibilités offertes par les technologies d'IA à l'échelle mondiale, et notamment aux économies émergentes et économies en développement. Ils ont souligné également que l'IA générative pouvait aider la société à relever certains défis urgents, comme l'amélioration des soins de santé et la recherche d'une issue à la crise climatique, et favoriser l'obtention de progrès au regard des Objectifs de développement durable (ODD).

Toutefois, les avantages potentiels de l'IA générative comportent aussi des risques

Le risque de voir s'aggraver les problèmes de désinformation et de manipulation de l'opinion est considéré par les membres du G7 comme l'une des principales menaces présentées par l'IA générative, au même titre que les risques d'atteintes aux droits de propriété intellectuelle et à la vie privée. D'après les premières tentatives de recensement, un millier environ d'incidents ou de risques différents, en lien avec l'IA générative, auraient été signalés dans quelque 5 600 articles de presse publiés entre janvier et juillet 2023.

Ces risques évoluant rapidement, les gouvernements des pays de G7 ont fait de leur résolution une priorité.

L'utilisation responsable de l'IA générative, la lutte contre la désinformation, la protection des droits de propriété intellectuelle et l'encadrement de l'IA générative figurent parmi les plus hautes priorités des responsables du G7 et exigent une coopération internationale entre partenaires attachés aux mêmes principes. Parmi les autres questions urgentes et importantes mises en avant par les membres du G7 figurent : la protection de la vie privée et la gouvernance des données, la transparence, l'équité et les biais, les droits de la personne humaine et les droits fondamentaux, la sécurité et la robustesse des systèmes d'IA et les effets de ces derniers sur le fonctionnement de la démocratie.

Les pays du G7 évaluent leur propre conduite face à l'IA générative, ainsi que les lacunes de leur cadre d'action.

Les pays comptent s'appuyer sur les cadres juridiques, existants et à venir, et définissent des directives pour maîtriser les risques liés à l'IA générative. Des initiatives nationales se renforcent par ailleurs dans le but d'exploiter les possibilités offertes par cette dernière. Les nouvelles problématiques soulevées par l'IA générative concernent plus particulièrement certains secteurs, comme l'éducation, les médias et l'environnement de travail.

Les membres du G7 s'accordent sur la nécessité de fournir, en ce qui concerne l'IA générative, des outils propres à garantir la sécurité, à assurer un contrôle qualité, à pourvoir au renforcement des capacités requises et à faire grandir la confiance.

La sécurité, le contrôle qualité, le renforcement des capacités et la confiance sont considérés au nombre des principaux domaines les plus urgents et les plus importants dans lesquels les pays de G7 pourraient mener une action d'envergure internationale. L'ouverture d'un dialogue est par ailleurs considérée comme ce qu'il y a de plus *urgent* ; l'élaboration de codes de conduite non contraignants comme l'une des actions les plus *importantes*.

Table des matières

RÉSUMÉ.....	3
INTRODUCTION.....	6
1. EXPOSÉ DU CONTEXTE	7
1.1. Tendances de l'IA générative	7
1.2. Incidents et dangers liés à l'IA générative.....	11
2. L'IA GÉNÉRATIVE VUE PAR LE G7	12
2.1. Possibilités et risques pour les membres du G7.....	12
2.2. Priorités des membres du G7 au regard des principes fondés sur les valeurs.....	15
2.3. Approche internationale collective envisageable	24
2.4. Initiatives nationales et régionales des membres du G7	28
Références	40

GRAPHIQUES

Graphique 1.1. L'IA générative s'est rapidement imposée sur la scène publique	7
Graphique 1.2. Les avancées de la recherche en IA générative ont précédé l'envolée des investissements	9
Graphique 1.3. Le code source libre d'IA générative a connu une progression plus graduelle que les modèles libres	10
Graphique 1.4. Les signalements d'incidents et de dangers associés à l'IA dans les médias de bonne réputation ont considérablement augmenté	12
Graphique 2.1. Les cinq principales <i>possibilités</i> offertes par l'IA en vue de la réalisation des objectifs nationaux et régionaux	13
Graphique 2.2. Les cinq <i>risques</i> principaux de l'IA générative pour la réalisation des objectifs nationaux et régionaux	14
Graphique 2.3. Les priorités les plus <i>urgentes</i> en ce qui concerne l'IA générative parmi les cinq priorités mentionnées dans la déclaration des Chefs d'État et de gouvernement du G7	15
Graphique 2.4. Les priorités les plus <i>importantes</i> en ce qui concerne l'IA générative, parmi les cinq priorités mentionnées dans la déclaration des Chefs d'État et de gouvernement du G7	16
Graphique 2.5. Les priorités les plus <i>urgentes</i> en ce qui concerne l'IA générative : autres priorités mentionnées dans les Principes de l'OCDE sur l'IA	17
Graphique 2.6. Les priorités les plus <i>importantes</i> en ce qui concerne l'IA générative : autres priorités mentionnées dans les Principes de l'OCDE sur l'IA	18
Graphique 2.7. Lacunes des mesures en vigueur ou en préparation chez les membres du G7 pour répondre correctement aux enjeux de l'IA générative	19
Graphique 2.8. Types de mesures dont le G7 pourrait recommander la mise en œuvre sans tarder	26
Graphique 2.9. Mesures les plus importantes que le G7 pourrait recommander	28

TABLEAUX

Tableau 2.1. Aperçu des initiatives nationales et régionales visant à tirer parti des possibilités offertes par l'IA générative	29
Tableau 2.2. Aperçu des initiatives nationales et régionales visant à faire face aux risques que présente l'ia générative	34

INTRODUCTION

En avril 2023, le Japon a accueilli la Réunion des ministres du Numérique et des Technologies du G7 à Takasaki. Les ministres se sont entendus sur une Déclaration ministérielle par laquelle ils mettent en lumière l'importance de discussions internationales sur l'interopérabilité des cadres de gouvernance de l'IA et le recensement des possibilités et enjeux associés à l'IA générative.

Dans leur Déclaration d'avril 2023, les ministres du Numérique et des Technologies du G7 reconnaissent « *la nécessité de faire un état des lieux, à brève échéance, des possibilités et enjeux [des technologies d'IA générative] et de continuer à promouvoir la sécurité et la confiance tandis que ces technologies continuent d'évoluer* » et s'engagent par conséquent « *à organiser de nouvelles discussions, à l'échelle du G7, au sujet de l'IA générative, qui pourront porter sur des thèmes tels que la gouvernance, la protection des droits de propriété intellectuelle, y compris celle du droit d'auteur, la promotion de la transparence, la lutte contre la désinformation, y compris la manipulation de l'information par des puissances étrangères, et l'utilisation responsable de ces technologies* » (paragraphe 47).

Les échanges que les ministres ont tenus sur l'IA ont été repris par les Chefs d'État et de gouvernement des pays du G7 lors de leur Réunion au sommet qui a eu lieu au mois de mai, à Hiroshima. Les Chefs d'État et de gouvernement sont convenus de charger leurs ministres de définir un « Processus d'Hiroshima sur l'IA », dans le cadre duquel les membres du G7 puissent poursuivre leurs échanges sur l'IA générative d'une manière inclusive.

L'IA générative peut s'entendre comme un type de modèle d'IA spécialement conçu pour produire de nouveaux éléments de contenu numérique (texte, image, audio, vidéo, code logiciel, entre autres), notamment lorsque ce genre de modèle est utilisé dans des applications et les interfaces utilisateur correspondantes. Les systèmes d'IA générative sont généralement construits comme des systèmes d'apprentissage automatique entraînés à l'aide de vastes ensembles de données. Leur fonctionnement consiste à prédire les mots, pixels, formes d'ondes, points de données, etc. qui ressembleraient au modèle ayant servi à leur entraînement, souvent en réponse à des invites (ou *prompts*) (OCDE, 2023^[1]), (OCDE, À paraître^[2]).

À l'appui du Processus d'Hiroshima sur l'IA, lancé par les Chefs d'État et de gouvernement des pays du G7, le Japon a diffusé un questionnaire aux membres du G7 en juin 2023. Le but, avec ce questionnaire, était de faire le point sur les initiatives prises ou envisagées par les membres du G7 en rapport avec les principales possibilités et les principaux risques de l'IA générative. Le questionnaire était structuré en quatre parties étroitement liées, comme suit :

1. Détermination des possibilités et des risques
2. Priorités au regard des principes fondés sur les valeurs
3. Possibilité d'une approche collective d'envergure internationale
4. Initiatives nationales et régionales

Le questionnaire était constitué principalement de questions fermées, en réponse auxquelles les participants étaient invités à classer différentes propositions ou à faire un choix parmi celles-ci, par exemple à partir d'une liste de possibilités et d'enjeux associés à l'IA générative, de priorités parmi les cinq domaines mentionnés dans la Déclaration des Chefs d'État et de gouvernement du G7 et parmi les Principes de l'OCDE sur l'IA, et de différentes mesures de politique publique que les membres du G7 pourraient recommander. Les questions ouvertes ont permis aux participants de présenter des initiatives d'envergure nationale ou régionale adoptées par les membres du G7 et touchant à l'IA générative.

Le présent rapport comporte un exposé des progrès de l'IA générative tels qu'ils ressortent des données de l'Observatoire OCDE des politiques relatives à l'IA et du Système OCDE de surveillance des incidents liés à l'IA (section 1) et une analyse des réponses au questionnaire diffusé aux membres du G7 (section 2).

1. EXPOSÉ DU CONTEXTE

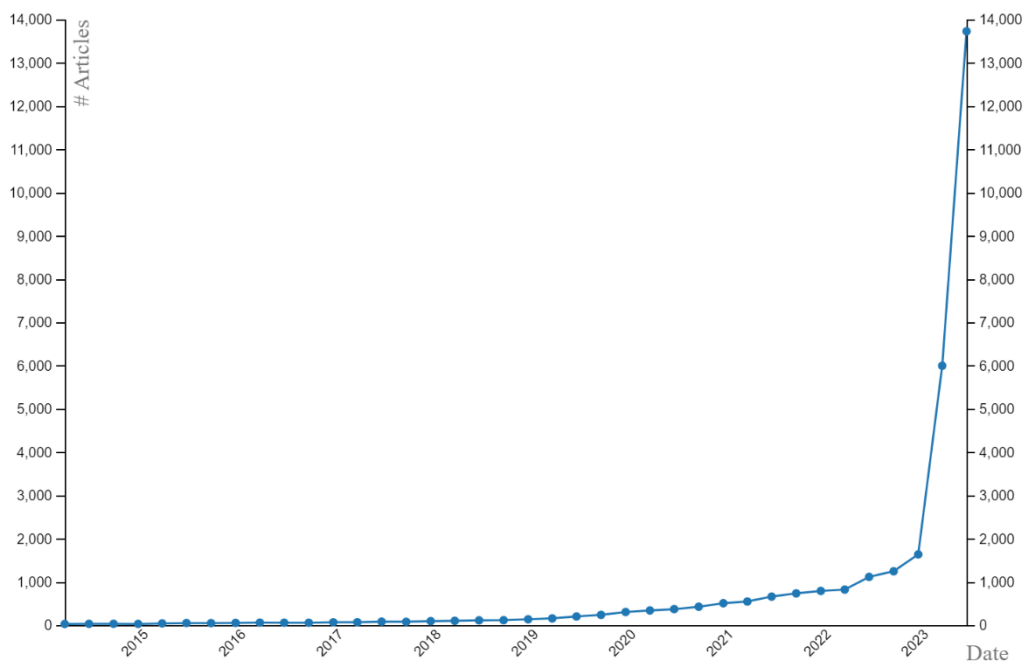
1.1. TENDANCES DE L'IA GÉNÉRATIVE

L'IA générative s'est imposée au cœur des discussions sur l'IA, parmi le grand public, dans les milieux universitaires et dans le monde politique. Elle laisse escompter la création d'une valeur économique considérable. Les entreprises ont commencé à adopter cette technologie pour s'ouvrir de nouveaux débouchés commerciaux, et des start-ups se disputent les investissements en capital-risque.

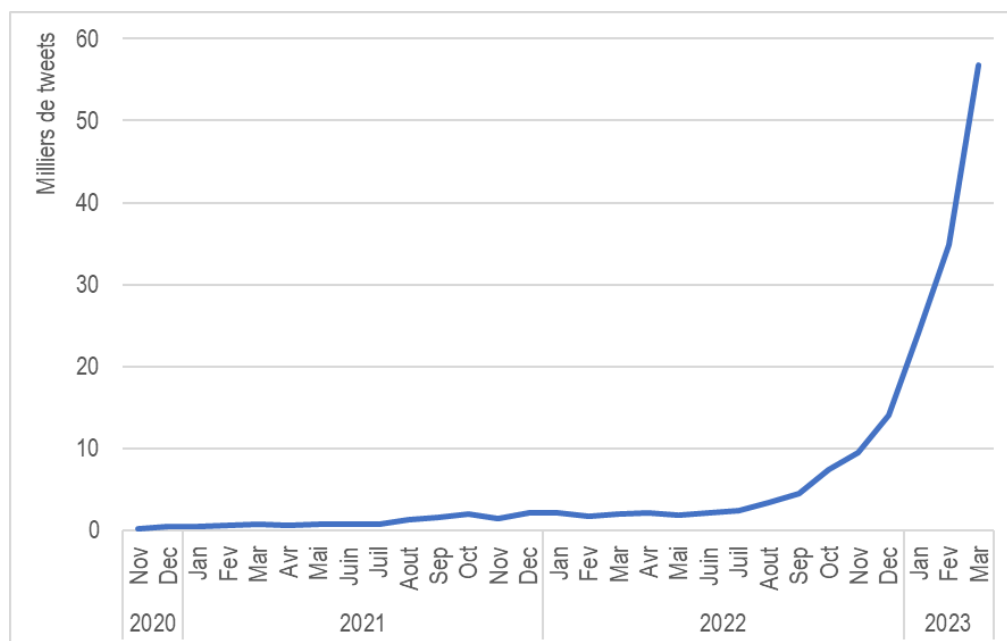
L'IA générative a fait son entrée sur la scène publique et devient un sujet de conversation chaque jour plus important, partout dans le monde, comme en témoigne la multiplication des articles de presse et des tweets qui lui sont consacrés. L'un et l'autre indicateurs ont été multipliés par huit en l'espace de six mois : le nombre d'articles de presse est ainsi passé de 1 600 au dernier trimestre 2022 à près de 14 000 au deuxième trimestre 2023, tandis que le nombre de tweets atteignait 57 000 en mars 2023, quand il n'était à l'origine que de 7 000, en octobre 2022 (Graphique 1.1, parties a et b, respectivement).

GRAPHIQUE 1.1. L'IA GÉNÉRATIVE S'EST RAPIDEMENT IMPOSÉE SUR LA SCÈNE PUBLIQUE

a) Nombre d'articles de presse sur l'IA générative, et autres thèmes connexes, publiés dans le monde



b) Nombre de tweets sur l'IA générative, et autres thèmes connexes, publiés dans le monde



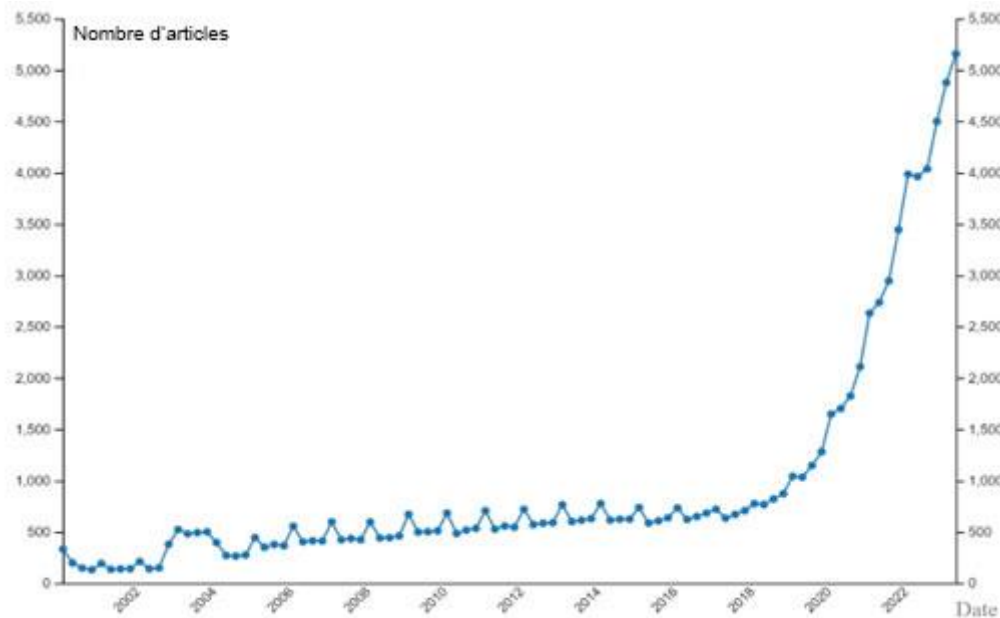
Note : Le nombre d'articles sur l'IA générative forme un sous-ensemble des articles consacrés à l'IA, qui comprennent les concepts wikidata suivants, ainsi que l'ensemble des concepts subalternes : IA générative, modèle transformateur et modèle génératif. Les tweets sur l'IA générative sont ceux contenant le terme *generative AI* ou le mot-dièse *#generativeAI*. Les données sur les articles de presse ont été ajustées à l'aide d'une technique, appelée lissage trimestriel, qui facilite les comparaisons et permet de mieux apprécier les tendances au fil du temps.

Source : OECD.AI, à partir de données d'*Event Registry*, pour les articles de presse, et de X, pour les tweets.

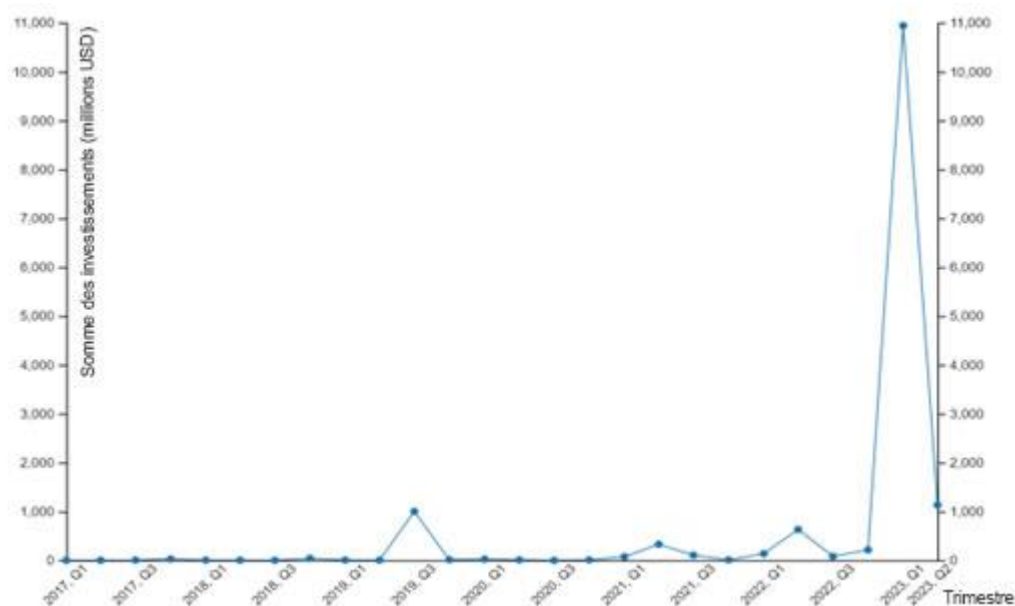
Les travaux de recherche et les investissements en capital-risque consacrés à l'IA générative se sont eux aussi sensiblement accrus. Le nombre de publications scientifiques touchant à l'IA générative a quintuplé depuis 2019, phénomène que l'on peut attribuer à l'intérêt grandissant porté aux innovations fondamentales, telles que les modèles transformateurs et les progrès de la puissance de calcul, et a ouvert la voie à la prolifération des grands modèles de langage (Graphique 1.2, partie a). Les investissements en capital-risque consacrés à l'IA générative au cours du premier semestre 2023 ont atteint un total de 12 milliards USD à l'échelle mondiale (Graphique 1.2, partie b). Les sommets atteints en 2019 et 2023 correspondent aux investissements de Microsoft dans OpenAI, pour un montant de 1 milliard USD et 10 milliards USD, respectivement.

GRAPHIQUE 1.2. LES AVANCÉES DE LA RECHERCHE EN IA GÉNÉRATIVE ONT PRÉCÉDÉ L'ENVOIÉE DES INVESTISSEMENTS

a) Nombre de publications scientifiques sur l'IA générative, et autres thèmes connexes, à l'échelle mondiale



b) Investissement total en capital-risque dans les start-ups spécialisées dans l'IA générative



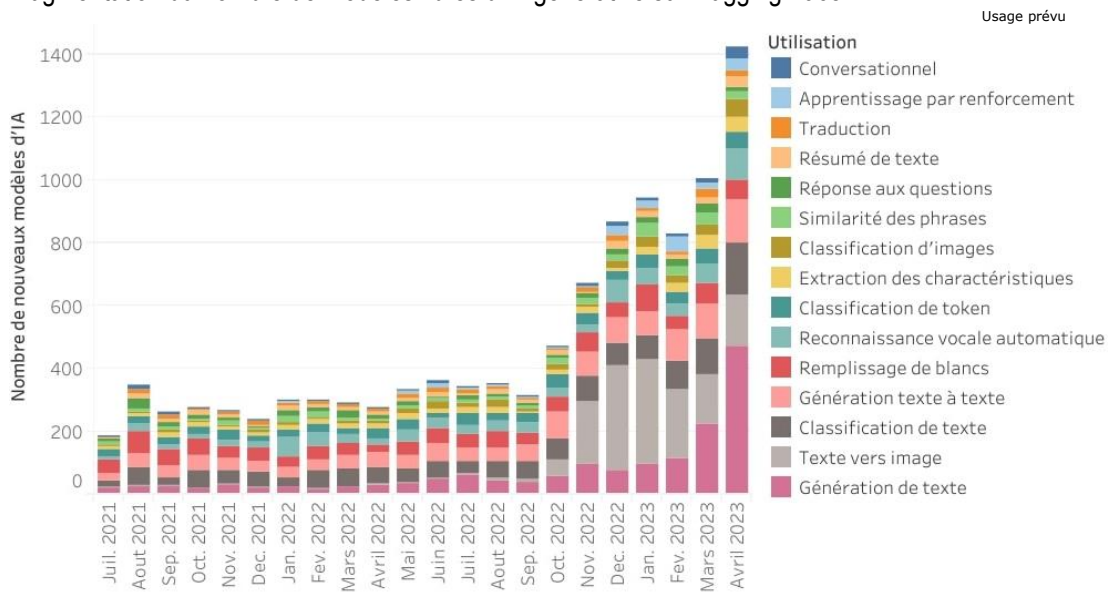
Note : Le nombre de publications scientifiques sur l'IA générative forme un sous-ensemble des publications consacrées à l'IA, qui comprend les concepts wikidata suivants, ainsi que l'ensemble des concepts subalternes : IA générative, modèle transformateur et modèle génératif. L'investissement en capital-risque consacré à l'IA générative est celui dont bénéficient des start-ups qui intègrent des notions telles que IA générative, réseau antagoniste génératif, génération de texte, générateur d'images, générateur de texte, générateur d'audio et modèle génératif dans leur description d'entreprise. Les deux ensembles de données font l'objet d'un lissage trimestriel destiné à en supprimer le bruit.

Source : OECD.AI, à partir de données d'OpenAlex, pour les publications scientifiques, et de Prequin, pour l'investissement en capital-risque.

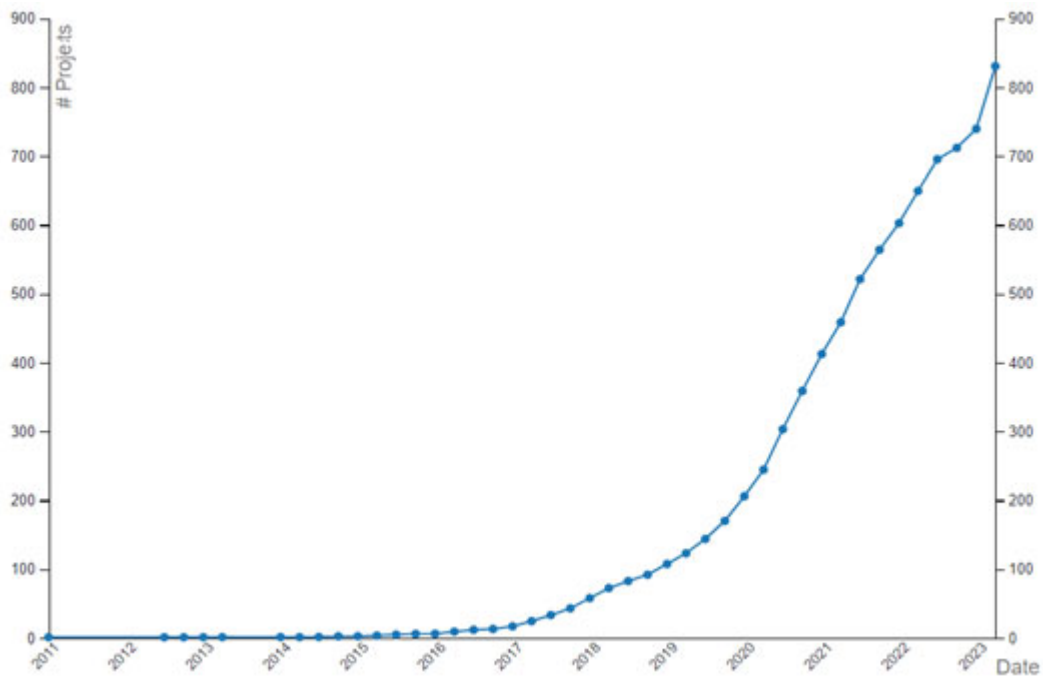
Les acteurs du logiciel libre ont toujours été le moteur des progrès de l'IA. Cette tendance se poursuit avec l'IA générative. Depuis le mois d'octobre 2022, la disponibilité et la création de modèles libres d'IA dédiés aux systèmes d'IA générative ont considérablement augmenté, comme en témoigne l'accroissement du nombre de modèles de génération de texte placés sur le référentiel Hugging Face au cours des derniers mois (Graphique 1.3, partie a). Par comparaison, le développement de code source libre d'IA sur la plateforme GitHub prend son essor à partir de 2017, et suit une progression plus graduelle. Ce panorama contrasté donne à penser que l'avènement de l'IA générative a été préparé par des avancées régulières et progressives dans le domaine du codage (Graphique 1.3, partie b).

GRAPHIQUE 1.3. LE CODE SOURCE LIBRE D'IA GÉNÉRATIVE A CONNU UNE PROGRESSION PLUS GRADUELLE QUE LES MODÈLES LIBRES

a) Augmentation du nombre de modèles libres d'IA générative sur Hugging Face



b) Progression de la production de code source libre sur GitHub



Note : Le code source libre d'IA générative forme un sous-ensemble de référentiels de GitHub consacrés à l'IA comprenant les concepts wikidata suivants, ainsi que l'ensemble des concepts subalternes : IA générative, modèle transformateur et modèle génératif. Les données de GitHub font l'objet d'un lissage trimestriel destiné à en supprimer le bruit.

Source : OECD.AI, à partir de données de Hugging Face, pour les modèles libres, et GitHub pour le code source libre.

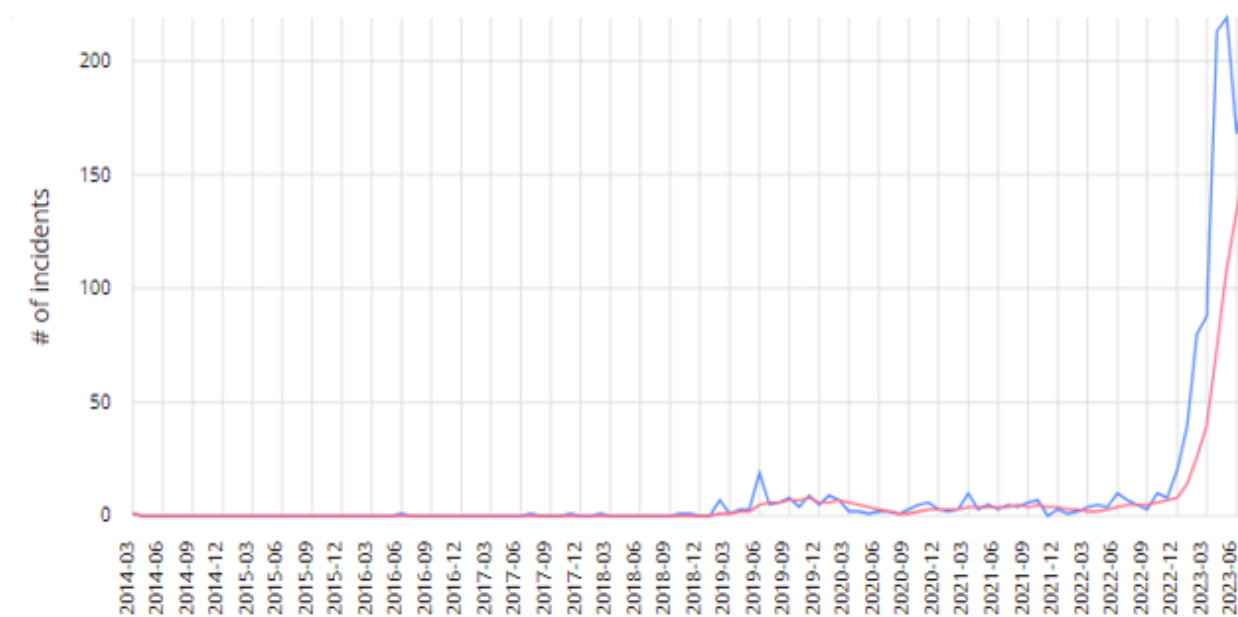
1.2. INCIDENTS ET DANGERS LIÉS À L'IA GÉNÉRATIVE

Si cette technologie est capable de révolutionner pour le meilleur l'industrie et la société, l'IA générative comporte également des risques pour les individus et la société elle-même. À titre d'exemple, elle peut être exploitée à des fins malveillantes ou criminelles, avec de graves conséquences à la clé, à travers, par exemple, la propagation de fausses informations et la création de contenus falsifiés (hypertrucage ou *deepfake*). La prise en compte de ce caractère ambivalent des technologies d'IA – dont l'IA générative – a incité l'OCDE à créer un Dispositif de surveillance des incidents liés à l'IA, conçu de manière à fournir des données en temps réel sur les risques liés à l'IA afin d'éclairer les décisions des pouvoirs publics. Pour ce faire, l'Organisation passe au crible les incidents et dangers concrets dont font état des médias de bonne réputation.

Le terme d'« incident » correspond au recueil d'un ou de plusieurs articles traitant d'un même événement¹. Entre janvier et juillet 2023, il a été signalé un millier environ d'incidents et de dangers associés à l'IA, dans quelque 5 600 articles de presse (Graphique 1.4). Si le Dispositif est encore en phase de développement (son lancement est prévu pour novembre 2023), les premières conclusions qu'il est possible d'en retirer mettent en lumière les risques inhérents aux systèmes d'IA générative et aident ainsi à façonner un paysage de l'IA plus sûr pour l'avenir.

¹Le Groupe d'experts OECD.AI sur les incidents liés à l'IA examine actuellement une définition de travail des incidents et dangers relatifs à l'IA. Aux fins du présent rapport, le terme « incident » est utilisé dans une acception large qui englobe également les dangers.

GRAPHIQUE 1.4. LES SIGNALEMENTS D'INCIDENTS ET DE DANGERS ASSOCIÉS À L'IA DANS LES MÉDIAS DE BONNE RÉPUTATION ONT CONSIDÉRABLEMENT AUGMENTÉ



Note : La ligne bleue correspond au nombre effectif d'incidents et de dangers signalés chaque mois. La ligne rouge correspond aux mêmes données, mais ajustées par lissage trimestriel. Le pic observé en 2019 vient d'une hausse des signalements en rapport avec l'hypertrucage.

Source : OECD.AI, Dispositif de surveillance des incidents liés à l'IA (à venir), à partir de données d'*Event Registry*.

2. L'IA GÉNÉRATIVE VUE PAR LE G7

Le présent rapport donne à voir les réponses reçues à un questionnaire conçu pour faciliter cet état des lieux et doit aider à éclairer l'examen, par le G7, des priorités d'action communes en ce qui concerne l'IA générative. La plupart des questions étant assorties d'une liste de propositions à classer ou à sélectionner, les classements proposés ne présentent pas les priorités par ordre absolu d'importance, mais un aperçu des réponses des pays au sujet des grandes priorités énoncées dans le questionnaire à un moment donné.

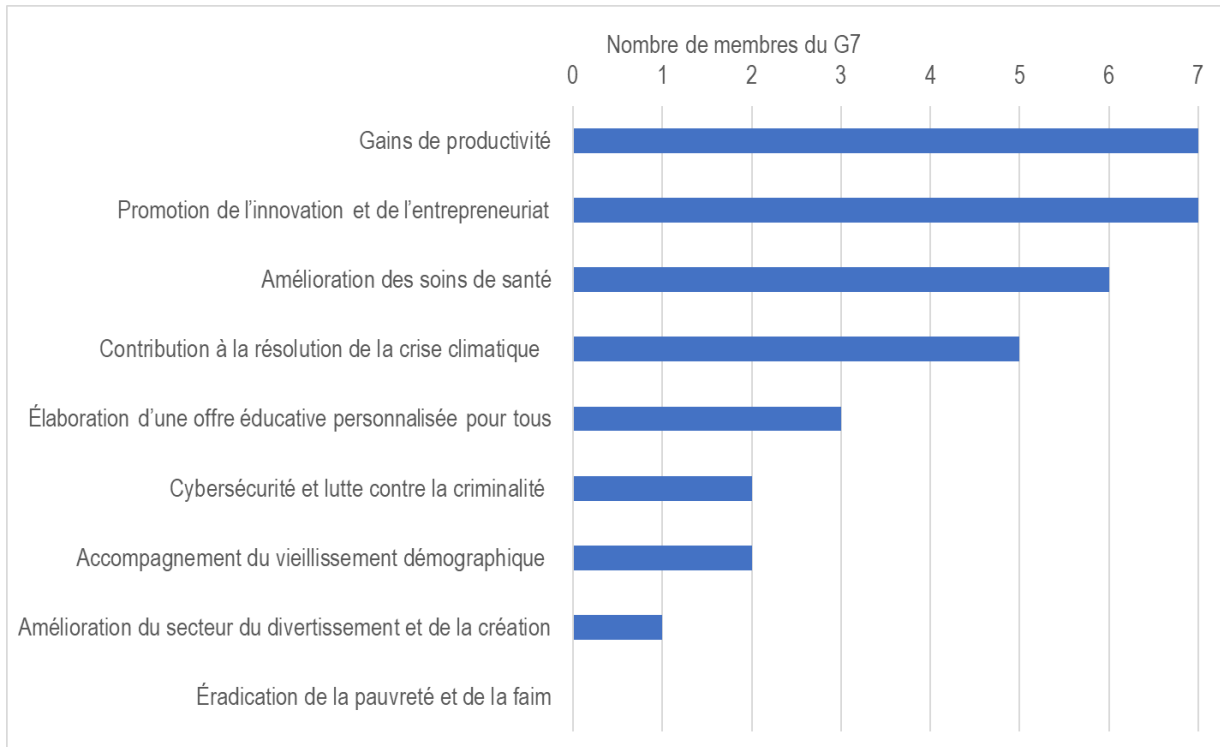
2.1. POSSIBILITÉS ET RISQUES POUR LES MEMBRES DU G7

Les gains de productivité et la promotion de l'innovation et de l'entrepreneuriat ont été considérés par l'ensemble des participants comme les principales possibilités offertes par l'IA générative parmi celles proposées dans le questionnaire. L'amélioration des soins de santé suivait de près, tout comme la résolution de la crise climatique (Graphique 2.1).

Le renforcement de la traçabilité et de la transparence des processus démocratiques et l'amélioration de l'accès aux services publics ont également été mentionnés au nombre des perspectives ouvertes par l'IA générative.

GRAPHIQUE 2.1. LES CINQ PRINCIPALES POSSIBILITÉS OFFERTES PAR L'IA EN VUE DE LA RÉALISATION DES OBJECTIFS NATIONAUX ET RÉGIONAUX

Nombre de membres du G7 ayant opté pour chaque possibilité (cinq choix possibles) parmi une liste préétablie



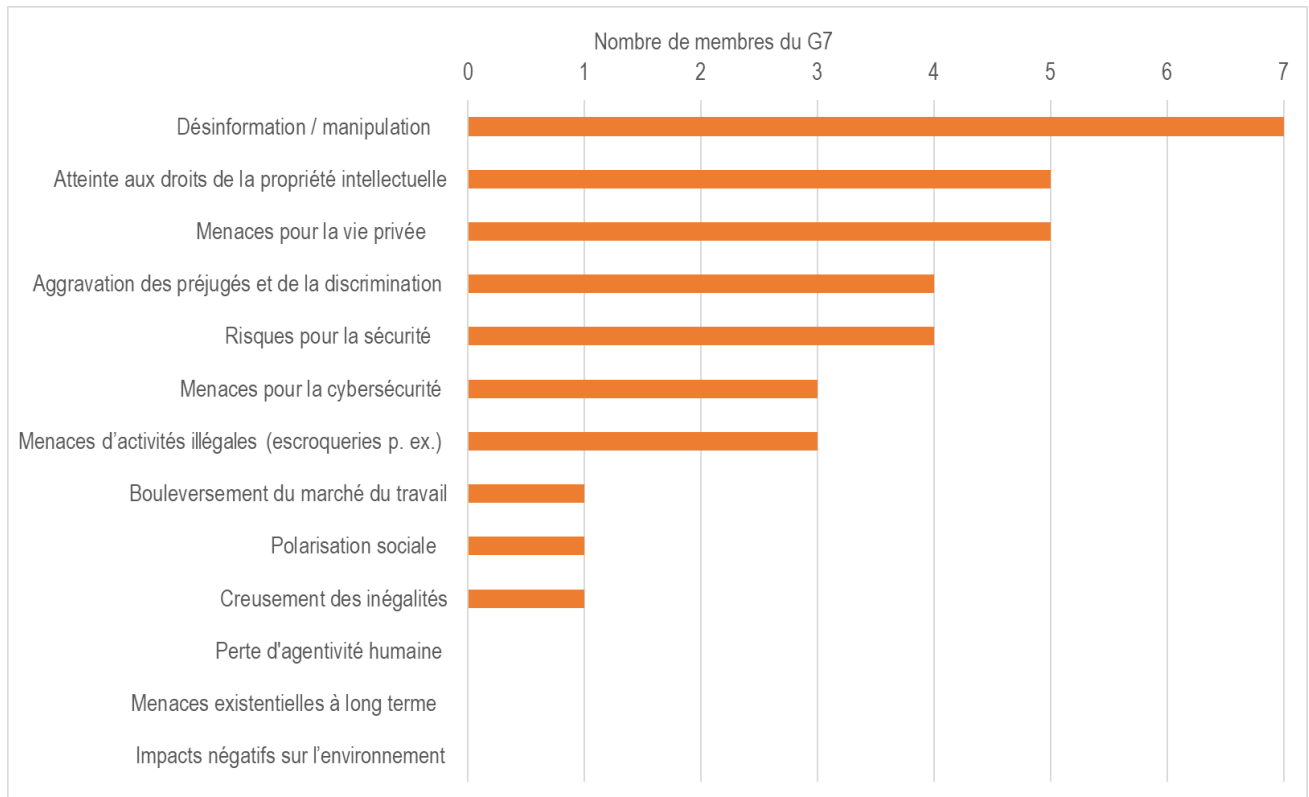
Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « Quelles sont, pour votre pays ou votre région, les cinq principales possibilités offertes par l'IA générative en vue de la réalisation des objectifs nationaux et régionaux ? (Merci d'indiquer cinq choix) ».

La désinformation et la manipulation de l'opinion qui en découle ont été considérées, à l'unanimité des participants, comme le principal risque présenté par l'IA générative parmi tous ceux proposés dans le questionnaire. De nombreux membres du G7 voient également les atteintes aux droits de propriété intellectuelle et les menaces pour la vie privée comme des risques importants (Graphique 2.2).

Les menaces sur la sécurité (notamment sur la cybersécurité), la manipulation et l'utilisation abusive des données, et les menaces sur les droits de la personne humaine ont également été mentionnées parmi les risques encourus.

GRAPHIQUE 2.2. LES CINQ RISQUES PRINCIPAUX DE L'IA GÉNÉRATIVE POUR LA RÉALISATION DES OBJECTIFS NATIONAUX ET RÉGIONAUX

Nombre de membres du G7 ayant opté pour chaque risque (cinq choix possibles) parmi une liste préétablie



Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « Quels sont, pour votre pays ou votre région, les cinq risques principaux de l'IA pour la réalisation des objectifs nationaux et régionaux ? (Merci d'indiquer cinq choix) ».

2.2. PRIORITÉS DES MEMBRES DU G7 AU REGARD DES PRINCIPES FONDÉS SUR LES VALEURS

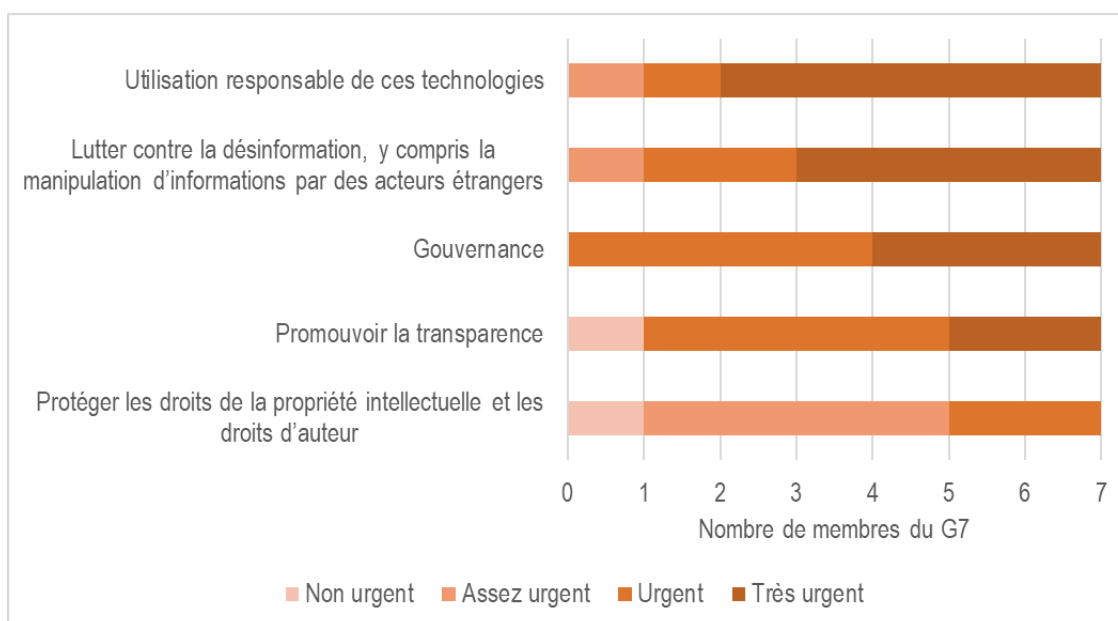
PRIORITÉS LES PLUS *URGENTES* ET LES PLUS *IMPORTANTES* DES MEMBRES DU G7 EN CE QUI CONCERNE L'IA GÉNÉRATIVE

L'utilisation « responsable » des technologies d'IA générative a été considérée par beaucoup comme la priorité la plus « urgente » pour les pouvoirs publics, parmi celles mentionnées dans la déclaration du G7. Suivaient de près la lutte contre la désinformation et l'encadrement adéquat de l'IA générative (Graphique 2.3).

Les menaces qui pèsent sur la cybersécurité et la biosécurité ont également été citées parmi les priorités les plus urgentes en ce qui concerne l'IA générative.

GRAPHIQUE 2.3. LES PRIORITÉS LES PLUS *URGENTES* EN CE QUI CONCERNE L'IA GÉNÉRATIVE PARMIS LES CINQ PRIORITÉS MENTIONNÉES DANS LA DÉCLARATION DES CHEFS D'ÉTAT ET DE GOUVERNEMENT DU G7

Nombre de membres du G7 ayant désigné chaque priorité comme urgente à partir d'une liste préétablie



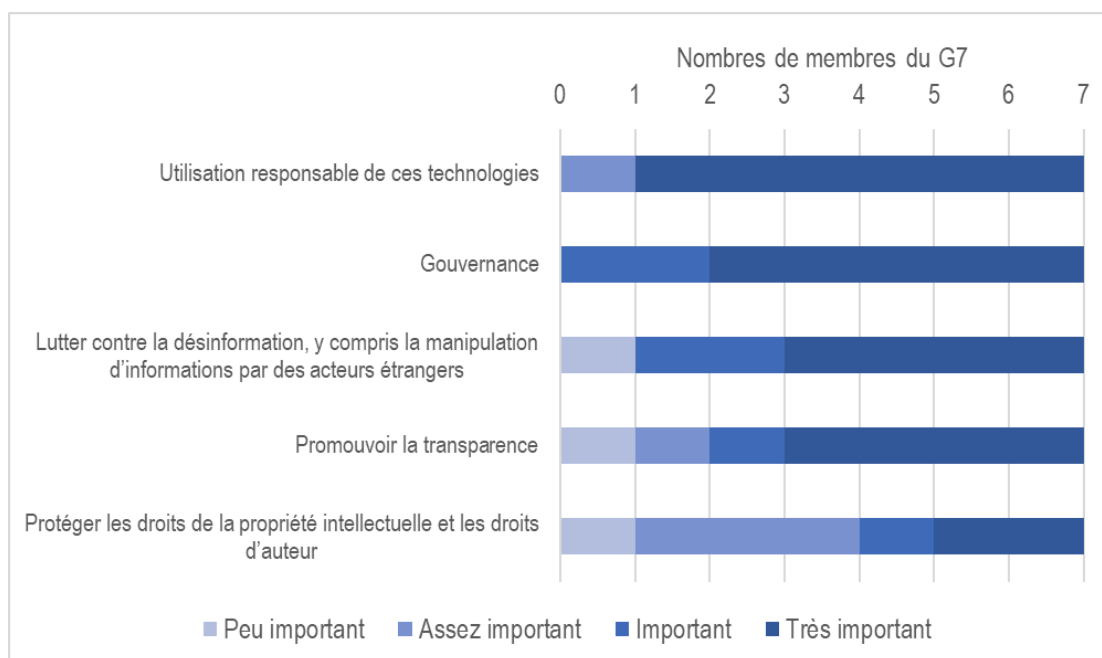
Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « D'un point de vue stratégique, quelles sont pour vous les priorités les plus urgentes et les plus importantes en ce qui concerne l'IA générative ? (Merci de classer les notions suivantes par ordre d'urgence et d'importance. Un même degré de priorité peut être attribué à plusieurs notions) ».

L'utilisation « responsable » des technologies d'IA générative a aussi été considérée comme la priorité la plus « importante » pour les pouvoirs publics, suivie par la gouvernance et la lutte contre la désinformation. Si les classements par ordre « d'importance » et « d'urgence » diffèrent légèrement, ce sont dans l'ensemble les mêmes priorités qui en ressortent (Graphique 2.4).

La menace qui pèse sur la cybersécurité a aussi été signalée comme une priorité importante, et celle pour la biosécurité comme l'une des plus hautes priorités dans le domaine de l'IA générative.

GRAPHIQUE 2.4. LES PRIORITÉS LES PLUS IMPORTANTES EN CE QUI CONCERNE L'IA GÉNÉRATIVE, PARMIS LES CINQ PRIORITÉS MENTIONNÉES DANS LA DÉCLARATION DES CHEFS D'ÉTAT ET DE GOUVERNEMENT DU G7

Nombre de membres du G7 ayant désigné chaque priorité comme importante à partir d'une liste préétablie

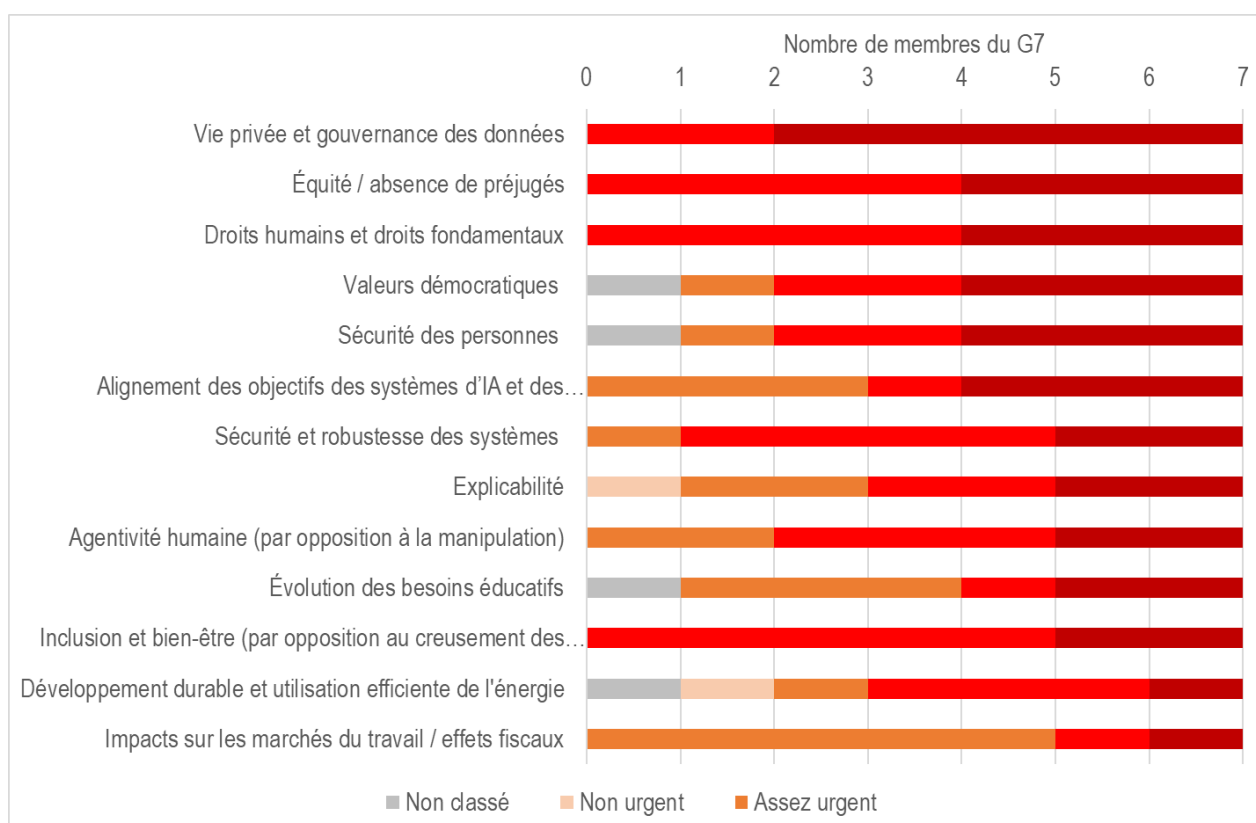


Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « D'un point de vue stratégique, quelles sont pour vous les priorités les plus urgentes et les plus importantes en ce qui concerne l'IA générative ? (Merci de classer les notions suivantes par ordre d'urgence et d'importance. Un même degré de priorité peut être attribué à plusieurs notions) ».

En outre, la protection de la vie privée et la gouvernance des données ont été signalées comme les plus urgentes des « autres » problèmes à traiter mentionnés dans le questionnaire, suivis par l'équité et les biais et par les droits de la personne humaine et les droits fondamentaux (Graphique 2.5).

GRAPHIQUE 2.5. LES PRIORITÉS LES PLUS URGENTES EN CE QUI CONCERNE L'IA GÉNÉRATIVE : AUTRES PRIORITÉS MENTIONNÉES DANS LES PRINCIPES DE L'OCDE SUR L'IA

Nombre de membres du G7 ayant désigné chaque priorité comme urgente à partir d'une liste préétablie

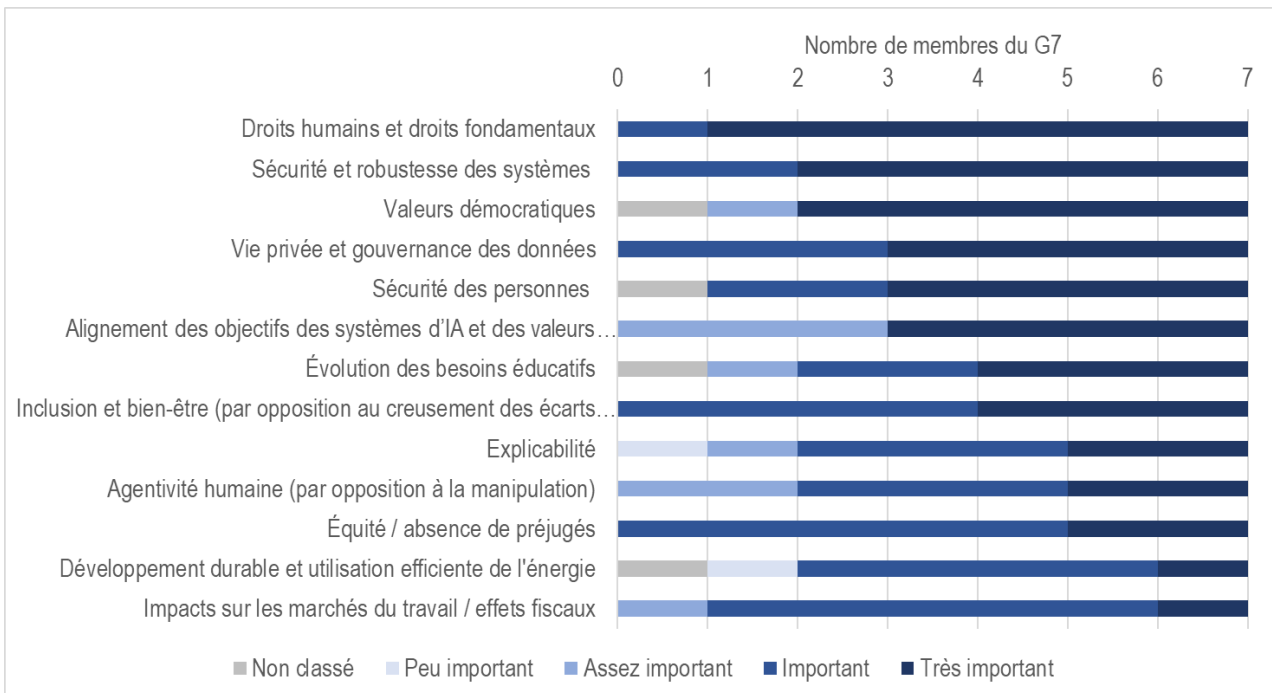


Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « D'un point de vue stratégique, quelles sont pour vous les priorités les plus urgentes et les plus importantes en ce qui concerne l'IA générative ? (Merci de classer les notions suivantes par ordre d'urgence et d'importance. Un même degré de priorité peut être attribué à plusieurs notions) ».

Les droits de la personne humaine et les droits fondamentaux, la sécurité et la robustesse des systèmes d'IA, les valeurs démocratiques, ainsi que la protection de la vie privée et la gouvernance des données, ont été considérés comme les plus importantes des « autres » priorités mentionnées dans le questionnaire (Graphique 2.6).

GRAPHIQUE 2.6. LES PRIORITÉS LES PLUS IMPORTANTES EN CE QUI CONCERNE L'IA GÉNÉRATIVE : AUTRES PRIORITÉS MENTIONNÉES DANS LES PRINCIPES DE L'OCDE SUR L'IA

Nombre de membres du G7 ayant désigné chaque priorité comme importante à partir d'une liste préétablie



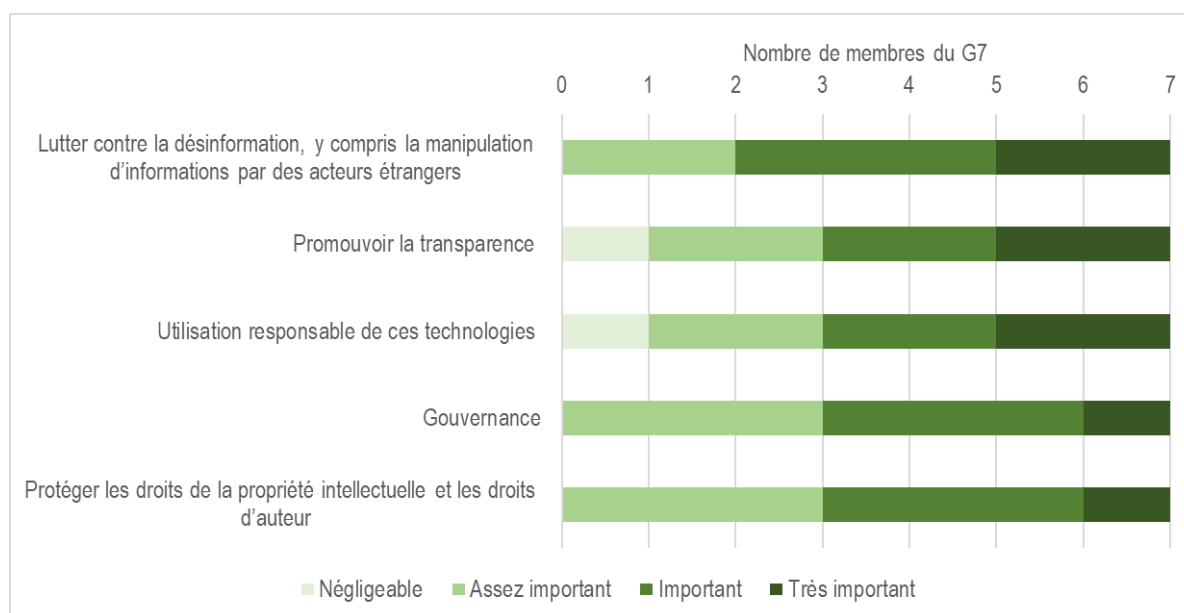
Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « D'un point de vue stratégique, quelles sont pour vous les priorités les plus urgentes et les plus importantes en ce qui concerne l'IA générative ? (Merci de classer les notions suivantes par ordre d'urgence et d'importance. Un même degré de priorité peut être attribué à plusieurs notions) ».

LACUNES LES PLUS SIGNIFICATIVES DES MESURES EN VIGUEUR OU EN PRÉPARATION EN RÉPONSE AUX ENJEUX DE L'IA GÉNÉRATIVE DANS LES PAYS DU G7

Les lacunes perçues par les membres du G7 sont très différentes ; plusieurs ont cité la désinformation, la transparence et l'utilisation responsable de l'IA comme les domaines où elles sont les plus nettes (Graphique 2.7).

GRAPHIQUE 2.7. LACUNES DES MESURES EN VIGUEUR OU EN PRÉPARATION CHEZ LES MEMBRES DU G7 POUR RÉPONDRE CORRECTEMENT AUX ENJEUX DE L'IA GÉNÉRATIVE

Lacunes auto-déclarées en lien avec les enjeux présentés



Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « De quelle ampleur sont les lacunes que vous observez dans les mesures (en vigueur ou en préparation) destinées à répondre correctement aux enjeux ci-après de l'IA générative ? ».

MESURES PRISES PAR LES MEMBRES DU G7 POUR COMBLER LES LACUNES CONNUES DE L'ACTION PUBLIQUE

Mise en application des cadres juridiques existants (ou à venir) et évaluation des lacunes de l'action publique

Les pays interrogés misent sur les cadres juridiques, existants ou à venir. Ils s'efforcent également de mesurer l'ampleur et la gravité des lacunes associées aux enjeux de l'IA générative.

- Le **Canada**, dans le cadre du projet de loi sur l'IA et les données (LIAD), propose une réglementation fondée sur les risques s'appliquant à la conception, au développement et à l'utilisation de systèmes d'IA – y compris d'IA générative – dans le secteur privé.
- La **France** étudie les nouveaux enjeux de l'IA générative au regard de la législation en vigueur (comme le RGPD et la Directive européenne relative au droit d'auteur) et précise que la législation européenne sur l'IA devrait couvrir quelques-unes des thématiques citées, dont la gouvernance et l'utilisation responsable des systèmes d'IA.
- L'**Allemagne** cherche à obtenir une plus grande cohérence stratégique entre les différents organismes publics et pour la définition de procédures pertinentes.
- En **Italie**, suite à l'avènement de l'IA générative (avec des technologies comme ChatGPT, par exemple), à l'accord provisoire du Parlement européen relatif à la législation sur l'IA et à l'accentuation des préoccupations et des risques éthiques associés à l'IA générative, début juillet 2023, le département de la Transformation numérique – rattaché à la Présidence du Conseil des ministres – a présenté une proposition de révision du « Plan stratégique de 2021 en faveur de l'intelligence artificielle ». La révision proposée est sur le point d'aboutir et fera l'objet d'une

enquête publique à partir du 30 septembre. Le « Plan stratégique en faveur de l'IA », tel que révisé, devrait entrer en vigueur le 31 décembre 2023. En outre, le gouvernement a institué un Comité permanent sur l'IA, relevant du Comité interministériel sur la transformation numérique, qui réunit des experts issus du milieu universitaire, de centres de recherche et d'associations d'entreprises italiennes. En mars 2023, l'Autorité de protection des données a imposé, avec effet immédiat, une limitation temporaire au traitement des données d'utilisateurs italiens par OpenAI. Dans sa décision, elle a souligné qu'aucune information n'était communiquée aux utilisateurs et personnes concernées quant à la nature des données collectées par OpenAI ; qui plus est, indiquait-elle, la collecte et le traitement de grands volumes de données à caractère personnel aux fins de l'entraînement des algorithmes utilisés par la plateforme ne reposaient sur aucune base juridique. À cela s'ajoute qu'aucune vérification de l'âge n'était effectuée pour s'assurer que les utilisateurs n'étaient pas mineurs.

- Le **Japon** a fait savoir qu'il applique le droit en vigueur (à travers le Code pénal et l'appareil judiciaire, par exemple) et les directives existantes (sur la R-D dans le domaine de l'IA, sur l'utilisation de l'IA, sur l'application des principes relatifs à l'IA, etc.), quoique les nouvelles problématiques soulevées par l'IA générative puissent en montrer les limites. L'émergence de l'IA générative faisant naître de nouveaux enjeux, il envisage de réviser ces directives et de les réunir en une seule, qui soit facile à comprendre par les développeurs, les fournisseurs, les utilisateurs et les autres entreprises.
- Au **Royaume-Uni**, c'est l'ensemble de l'administration qui travaille de concert à évaluer l'ampleur et la gravité des lacunes des mesures d'atténuation en vigueur, et l'on y étudie de nouvelles mesures pour chaque maillon de la chaîne logistique de l'IA.
- De même, aux **États-Unis**, en vertu d'un décret exigeant le respect de certains principes dans le cadre des activités fédérales en rapport avec l'IA, le Bureau de la gestion et du budget met actuellement au point des orientations établissant des règles que les agences et départements fédéraux devront suivre pour renforcer la gouvernance de l'IA, promouvoir les achats publics de systèmes d'IA et gérer le risque algorithmique afin de protéger les droits et la sécurité du peuple américain. L'Institut national de normalisation et de développement technologique (NIST), qui relève du Département du Commerce, a publié un cadre de gestion des risques liés à l'IA (AI RMF) en janvier 2023 et constitué, en juin suivant, un groupe de travail public sur l'IA générative, chargé de définir un cadre de gestion des risques adapté à cette dernière.
- Le projet de **législation européenne sur l'IA** vise à promouvoir le développement et l'adoption de l'IA tout en maîtrisant les risques que certains systèmes, dont les systèmes d'IA générative, peuvent présenter pour la sécurité et les droits fondamentaux. En outre, l'UE compte renforcer son arsenal réglementaire par la révision de la législation en vigueur, avec, par exemple, l'adoption récente du règlement sur les machines, ainsi que le projet (en cours d'examen) de révision de la directive sur la responsabilité du fait de produits défectueux et le projet de loi sur la cyberrésilience. À côté des mesures réglementaires, l'UE applique un Plan coordonné sur l'IA avec ses États membres et travaille au Pacte sur l'IA. Ce Pacte doit encourager les entreprises à faire connaître de leur plein gré les processus et les pratiques qu'elles mettent en place pour préparer leur mise en conformité à la législation de l'UE sur l'IA et faire en sorte que la conception, le développement et l'utilisation des systèmes d'IA soient dignes de confiance.

Élaboration de lignes directrices, création de nouvelles orientations et mise en place d'organes de gouvernance

Le Secrétariat du Conseil du Trésor (SCT) du **Canada** envisage de publier des lignes directrices sur l'utilisation de l'IA générative au sein du gouvernement fédéral. Ce « guide » procurera aux institutions fédérales des orientations relatives à l'utilisation de ces outils. Il donne une vue d'ensemble de l'IA générative, identifie les enjeux et les préoccupations liés à son usage, présente des principes pour une utilisation

responsable, et expose des considérations intéressantes l'action publique et des bonnes pratiques en la matière. En outre, début août, le Canada a lancé des tables rondes afin de recueillir les observations des parties prenantes concernant une proposition de code de pratique canadien sur l'IA générative. Ce code fournira des orientations volontaires aux entreprises qui développent et utilisent des systèmes d'IA générative, et les aidera à préparer leurs processus et leurs produits avant que la réglementation formelle prenne effet.

De même, le Bureau de la gestion et du budget des **États-Unis** met actuellement au point des orientations établissant des règles que les agences et ministères fédéraux devront suivre pour renforcer la gouvernance de l'IA, promouvoir les achats publics de systèmes d'IA et gérer le risque algorithmique afin de protéger les droits et la sécurité du peuple américain.

L'**Allemagne** indique avoir pris plusieurs mesures spécifiques, avec notamment : a) la mise en place d'un centre consultatif pour l'utilisation de l'IA dans le secteur public, qui mènera des activités de renforcement des capacités et de réseautage ; et b) la création d'un Centre sur la qualité et l'innovation en matière d'IA.

L'**Union européenne** envisagerait d'inscrire dans sa législation sur l'AI une structure de gouvernance et des règles spécifiques visant à garantir que la conception, le développement et l'utilisation des technologies d'IA, y compris de l'IA générative, soient dignes de confiance.

Nécessité d'une gouvernance internationale

Lorsque le recours aux systèmes juridiques existants ne permet pas de remédier aux lacunes actuelles, les membres du G7 jugent nécessaire de réfléchir à d'autres mesures d'atténuation, de s'inspirer de l'expérience d'autres pays, et de coordonner les mesures au niveau international afin d'élaborer des approches globales et cohérentes entre les pays. Un pays membre du G7 a fait référence aux mécanismes internationaux qui visent à lutter contre la désinformation provenant de sources étrangères, tels que le Centre de ressources de l'OCDE sur la désinformation et la mésinformation, le Mécanisme de réponse rapide du G7 et la Cohorte pour l'intégrité de l'information du Sommet pour la démocratie.

EXEMPLES DE DOMAINES ET DE SECTEURS DANS LESQUELS LES LOIS ET POLITIQUES EXISTANTES SONT APPLIQUÉES AUX NOUVELLES PROBLÉMATIQUES INDUITES PAR L'IA GÉNÉRATIVE OU SONT REMISES EN CAUSE PAR CELLES-CI

Les lois et politiques existantes, y compris celles relatives à la protection des consommateurs et de la vie privée, s'appliquent à l'IA générative. Plusieurs membres du G7 ont signalé que les législations propres à différents secteurs sont applicables à l'IA générative, mais qu'elles sont également remises en question par celle-ci. Ils ont notamment souligné les défis soulevés en matière de respect de la vie privée et des droits de la propriété intellectuelle, y compris du droit d'auteur. Certains pays ont indiqué que l'IA entre déjà dans le périmètre couvert par la législation sur la protection des données ou par les régimes de protection de la propriété intellectuelle existants. D'autres ont fait valoir que l'IA soulève de nouvelles questions juridiques concernant la propriété du contenu créé en tout ou partie avec l'IA, comme les images et les textes, ainsi que sur la façon dont les droits associés aux données d'entraînement influent sur le statut juridique des résultats des modèles, ajoutant que ces questions sont actuellement étudiées au niveau national.

Les membres du G7 ont également cité des exemples de secteurs particulièrement menacés par l'IA générative. Il s'agit des secteurs de la création, du travail intellectuel, du droit, de la cybersécurité, de la santé et des technologies médicales, du secteur financier et des services publics fédéraux.

D'après les réponses de plusieurs membres du G7, les problèmes posés par l'IA générative semblent être particulièrement aigus dans les secteurs suivants :

- **Éducation** : Plusieurs pays membres du G7 ont signalé que le secteur de l'éducation mérite une attention particulière dans la mesure où il est déjà touché par l'IA générative, et que les conséquences

devraient s'amplifier à l'avenir. Un membre du G7 réunit des experts afin de travailler avec le secteur de l'éducation en vue de recenser et de partager les meilleures pratiques et les possibilités offertes par l'IA.

- **Environnement de travail** : Deux membres du G7 ont souligné que l'IA générative a une incidence sur les questions liées à l'environnement de travail. L'IA générative peut accélérer les processus de recrutement (au moyen d'agents conversationnels pilotés par l'IA générative, par exemple), mais si les données sont biaisées, l'équité des processus de recrutement peut s'en trouver compromise. En outre, ils ont fait observer qu'au travail, les employés peuvent utiliser l'IA générative sans instructions ou réglementation appropriées, et divulguer des données d'entreprise et/ou des informations à caractère personnel sensibles ou confidentielles à des tiers extérieurs à l'entreprise.
- **Communication et médias** : Deux membres du G7 ont indiqué que le journalisme et l'information sont d'autres domaines dans lesquels les lois et politiques existantes sont appliquées à l'IA générative ou sont remises en cause par celle-ci, soulignant le risque que l'IA générative serve à la production d'informations erronées et d'hypertrucages.

CARACTÉRISTIQUES SPÉCIFIQUES DE L'IA GÉNÉRATIVE QUI REMETTENT EN QUESTION LA RÉGLEMENTATION/GOUVERNANCE

Imprévisibilité, adaptabilité, autonomie et nature polyvalente de l'IA générative

Plusieurs membres du G7 considèrent que l'imprévisibilité des effets, l'adaptabilité, l'autonomie et la polyvalence sont des caractéristiques de l'IA générative qui en compliquent la réglementation et la gouvernance.

- L'IA générative est à l'origine de nouvelles **innovations de rupture** qui touchent un large éventail d'utilisateurs et de développeurs inexpérimentés. Le vaste champ d'application de la technologie, les interactions croissantes entre les systèmes d'IA générative, ainsi que les progrès techniques rapides sont source à la fois d'une grande incertitude et d'une grande imprévisibilité pour la société.
- De nombreuses applications d'IA générative ont aujourd'hui peu d'**autonomie**, c'est-à-dire peu de capacité à prendre des décisions indépendantes, sans supervision ou direction humaine. Elles produisent souvent du contenu lorsqu'elles en reçoivent l'instruction à partir d'invites. Toutefois, les résultats de l'IA générative peuvent déclencher une action autonome ou partiellement autonome programmée dans un logiciel d'automatisation classique, les résultats de l'IA générative formant le point de départ de cette action. Pour l'avenir, toutefois, à mesure que les systèmes d'IA générative, à l'instar des grands modèles de langage, seront de plus en plus utilisés comme des « agents » génératifs autonomes dotés d'extensions permettant de les connecter à des applications tierces, ils gagneront en autonomie. Par exemple, les extensions permettent aux modèles de langage d'exploiter des données récentes, y compris des informations en temps réel, telles que les cours des actions ou des articles de presse, et d'aider les utilisateurs selon des modalités inédites (en facilitant les commandes et les réservations autonomes, etc.).
- Du fait de « l'**adaptabilité** » de l'IA, il peut être difficile pour les développeurs de comprendre l'intention ou la logique qui aboutit aux résultats du système. En effet, les systèmes d'IA sont « entraînés » en déduisant des schémas et des connexions dans les données que les programmeurs humains ont souvent du mal à discerner. Ce mécanisme permet aux systèmes d'IA générative d'acquérir la capacité d'exécuter de nouvelles tâches ou de mener des raisonnements auxquels les développeurs ne s'attendaient pas, ce qui offre un potentiel considérable, mais fait également obstacle à la conception intentionnelle ou même à la parfaite compréhension des capacités du

modèle. Ce décalage entre les connaissances et l'intention des développeurs et les capacités du système peut compliquer l'attribution de la responsabilité des résultats.

- L'adaptabilité et la nature polyvalente de l'IA générative, ainsi que sa capacité à développer et à diffuser des contenus beaucoup plus rapidement qu'auparavant, **peuvent exacerber les biais et autres risques** plus fréquemment que dans les systèmes antérieurs, en favorisant ou en renforçant les représentations stéréotypées ou préjudiciables.
- La diversité des cas d'utilisation et des contextes auxquels l'IA générative peut être appliquée pose également problème en soi, car chacun d'eux peut nécessiter un traitement réglementaire ou un régime de gouvernance différent.
- L'opacité, la complexité et l'adaptation continue des systèmes d'IA peuvent générer de nouveaux risques ou aggraver les risques existants pour la santé, la sécurité et les droits fondamentaux.

Manque de transparence

Le manque de transparence de l'IA générative, à la fois pendant la phase de développement (transparence des développeurs sur la manière dont ils ont développé le système) et celle d'utilisation (transparence des utilisateurs sur le fait qu'ils utilisent un système) a été évoqué par un membre du G7 comme une problématique qui complique la réglementation et la gouvernance. Un membre du G7 met actuellement au point une législation imposant des obligations de transparence, y compris pour certains systèmes d'IA générative, qui a été citée en exemple pour répondre à cette préoccupation.

Mésinformation et désinformation (y compris campagnes de désinformation depuis l'étranger)

Un membre du G7 a insisté sur la capacité des modèles d'IA générative à créer des contenus synthétiques (des hypertrucages, par exemple) à grande échelle avec peu de ressources ou d'expertise. Il a notamment fait observer que la prochaine génération de médias interactifs génératifs exploitera un contenu d'influence ciblé qui est hautement personnalisé, localisé et conversationnel. Un autre membre du G7 s'est dit préoccupé par la capacité du contenu généré par l'IA à influencer à grande échelle le comportement, l'expression et les émotions des individus, ainsi qu'à relayer ou à promouvoir des informations erronées. En outre, il a mis en garde contre les contenus erronés ou inventés présentés comme des faits (c'est-à-dire des résultats qui inspirent un faux sentiment de confiance ou « hallucinations »).

Un faible niveau de culture numérique peut accélérer encore la propagation de la mésinformation et aggraver son impact, ce qui nécessite d'améliorer la maîtrise du numérique et les outils d'évaluation de l'authenticité des contenus.

Les volumes de données considérables posent des problèmes de réglementation

Deux membres du G7 ont indiqué que l'entraînement de l'IA générative fait appel à des volumes considérables de données publiquement disponibles, qui peuvent être utilisées sans autorisation et peuvent donc enfreindre les lois sur la protection des données et sur le droit d'auteur. Aussi, les données d'entraînement soulèvent des problèmes de droits d'auteur lorsque les résultats ressemblent aux sources originales. Un autre membre du G7 a également relevé le manque de clarté concernant le type de données d'entraînement utilisées ainsi que les conséquences pour la protection des consommateurs, la protection de la propriété intellectuelle et l'application de la loi, en particulier compte tenu de la difficulté à empêcher les modèles de reproduire des contenus protégés par le droit d'auteur. Il a également souligné le fait que les systèmes de réglementation et de gouvernance peinent à suivre le rythme des progrès rapides des capacités de l'IA.

2.3. APPROCHE INTERNATIONALE COLLECTIVE ENVISAGEABLE

DÉFIS COMMUNS SOULEVÉS PAR L'IA GÉNÉRATIVE QUI NÉCESSITENT UNE COHÉRENCE ET UNE COLLABORATION À L'ÉCHELLE INTERNATIONALE

Cadres réglementaires et interopérabilité

Dans leurs réponses, les membres du G7 ont évoqué la nécessité de mettre en place une réglementation et une surveillance appropriées.

- Un membre du G7 a rappelé que les cadres existants, comme les Principes de l'OCDE sur l'IA, exigent des normes, des standards et des processus d'évaluation internationaux. Il a donc appelé de ses vœux une cohérence et une collaboration à l'échelle internationale, y compris avec les pays en développement.
- Un autre membre du G7 a souligné l'importance de la gouvernance internationale (par exemple, établir des cadres de gouvernance communs et des normes internationales sur la fiabilité de l'IA générative, c'est-à-dire le contrôle de la qualité) et la nécessité pour les institutions internationales de faciliter une action rapide et coordonnée entre les pays. L'objectif est de leur permettre de réagir à des menaces actuellement inconnues. Ce même membre du G7 a fait observer que différents domaines de la gouvernance mondiale – du changement climatique au commerce international – ont bénéficié de la codification et de l'institutionnalisation de la coopération entre les pays.
- Un autre membre du G7 a également souligné le besoin d'élaborer des principes précis et détaillés pour permettre leur mise en œuvre dans les pays du G7. Il a ajouté que ces principes pourraient être appliqués au moyen d'accords généraux avec les fournisseurs de systèmes d'IA générative ou, à défaut de tels accords, par l'adoption d'une législation contraignante.
- D'autres membres ont insisté sur l'impératif d'améliorer l'interopérabilité des cadres réglementaires dans les différents pays. Un autre membre du G7 a par ailleurs souligné la nécessité de lignes directrices communes pour promouvoir le développement et l'utilisation responsables de l'IA.

Risques

Plusieurs membres du G7 ont souligné l'importance d'empêcher que l'IA générative soit utilisée pour générer des menaces chimiques ou biologiques (des virus, par exemple) ou une désinformation/mésinformation massive (y compris émanant d'acteurs étrangers), d'où la nécessité d'une coopération internationale pour affronter ces défis communs. Les problèmes de sécurité constituent un autre défi soulevé par l'IA générative, dont la solution passe par un alignement et une collaboration à l'échelle internationale, justifiant un appel à traiter les risques internationaux de cybersécurité liés à l'IA au niveau mondial. Des membres ont également mentionné le risque de fragilisation de la stabilité sociale. D'autres encore ont signalé que les responsables de l'action publique ont du mal à suivre le rythme rapide des évolutions technologiques.

Données à caractère personnel et droits de propriété intellectuelle

Un membre du G7 a souligné la nécessité de prévoir des mesures relatives à l'utilisation des données à caractère personnel ainsi que des contenus protégés par des droits de propriété intellectuelle. Ces mesures devraient distinguer les données servant à l'entraînement des modèles de celles utilisées par le système pour interagir avec l'utilisateur final et générer du contenu. D'autres membres ont appelé à rechercher un terrain d'entente au niveau international sur la question de la gestion des données utilisées pour entraîner les systèmes d'IA. Un membre du G7 a suggéré d'établir une traçabilité des données fournies par les opérateurs humains afin qu'elles puissent être supprimées par le fournisseur d'origine.

Transparence

Plusieurs membres du G7 ont insisté sur le besoin de transparence en ce qui concerne le développement et l'utilisation des systèmes d'IA générative. Certains participants ont cité les travaux menés par l'Union européenne (la législation européenne sur l'IA) relatifs à l'articulation des responsabilités entre les acteurs des chaînes de valeur de l'IA. Ils ont notamment fait valoir que les fournisseurs de modèles de fondation devraient assurer une transparence suffisante afin que les fournisseurs des « produits » finals puissent les mettre sur le marché en toute sécurité. Ils ont aussi rappelé que le projet de législation de l'UE sur l'IA contient les dispositions suivantes en matière de transparence des systèmes d'IA générative : (a) un robot conversationnel opéré par l'IA générative serait soumis à des obligations de transparence ; (b) un système d'IA générative pouvant être utilisé directement dans des applications à haut risque (par exemple, pour l'évaluation de candidats) devrait satisfaire aux exigences applicables aux systèmes d'IA à haut risque, lesquelles comprennent des dispositions relatives à la transparence.

Éthique

Certains membres du G7 ont mentionné la nécessité de mettre en œuvre des politiques communes en matière d'éthique dans le domaine de l'IA afin de remédier aux biais présents dans les systèmes d'IA. Un membre a également fait valoir la nécessité de définir un code d'éthique afin de restreindre toute utilisation abusive de l'IA générative de manière générale, mais aussi pour encadrer son utilisation dans des contextes militaires.

L'intelligence artificielle au service de l'intérêt commun

Un membre du G7 a insisté sur la nécessité d'une collaboration internationale pour permettre une gestion responsable du déploiement et de l'utilisation de ces technologies par les pouvoirs publics au service de l'intérêt collectif dans divers secteurs, notamment celui de la santé (pour améliorer les diagnostics médicaux) et de l'administration publique (pour rendre son fonctionnement plus efficace et efficient). D'autres membres ont aussi déclaré que leur objectif était de promouvoir une utilisation de l'IA dans l'intérêt de tous. Un membre du G7 a souligné que les progrès rapides de l'IA générative ouvrent de nouvelles perspectives aux pays émergents et en développement, et que la coopération internationale visant à favoriser une innovation plus locale dans le domaine de l'IA devrait se concentrer sur le développement d'ensembles de données représentatifs, de compétences en IA, de cadres d'action pour une IA responsable et de régimes appropriés en matière de protection des données.

Inégalité des chances (au sein des sociétés et pour les pays du Sud)

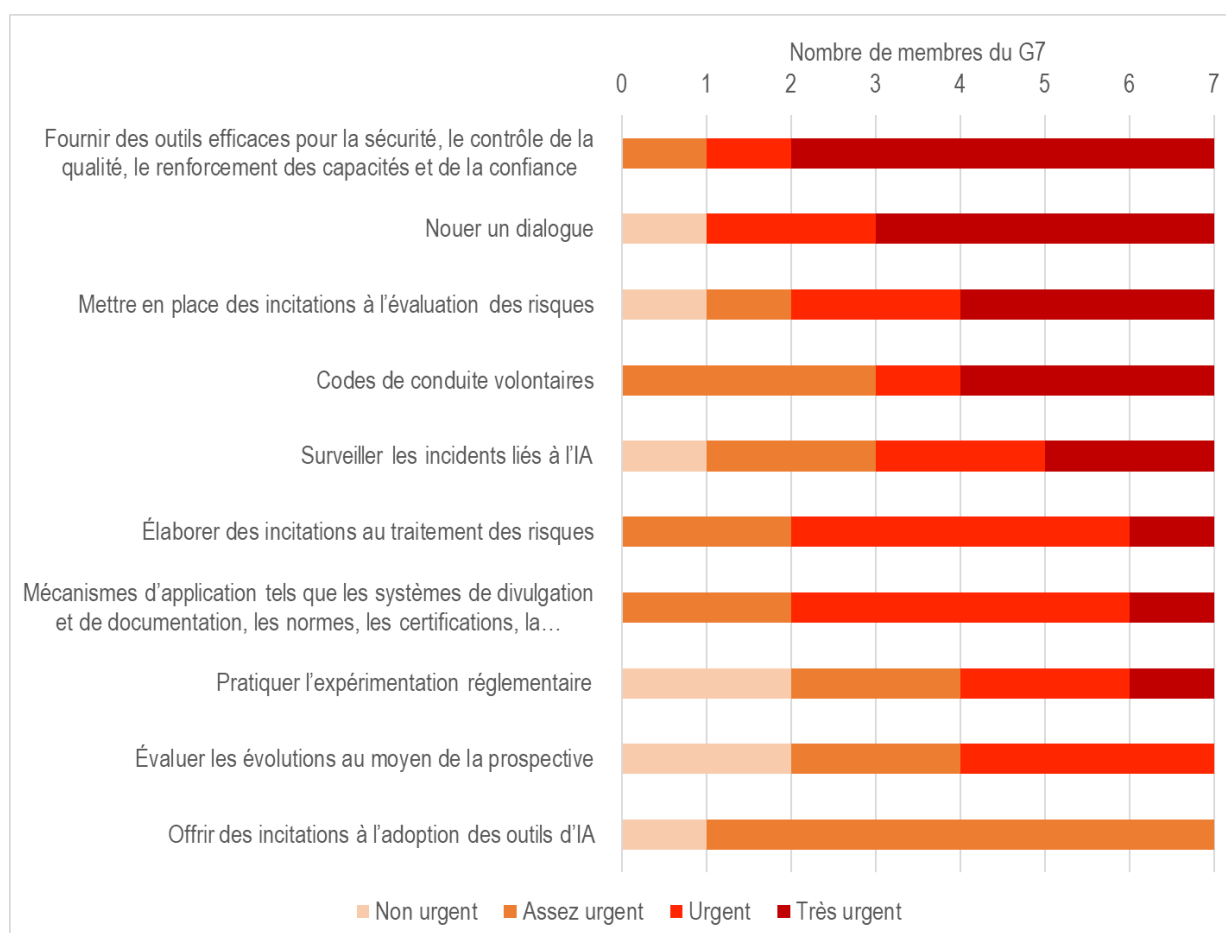
Un membre du G7 a fait observer que si les progrès rapides de l'IA générative ouvrent de nouvelles perspectives aux pays du Sud, ils présentent aussi certains risques. Il a notamment insisté sur le retard pris par les pays d'Asie et d'Afrique s'agissant de la mise en place de cadres d'action inclusifs favorisant une IA fondée sur des valeurs. L'IA générative peut également aggraver les fractures numériques ; les modèles étant entraînés à partir de données susceptibles de comporter des biais, cela risque en effet d'accentuer la polarisation à l'échelle mondiale. C'est pour cela qu'un membre du G7 a appelé à associer à la gouvernance de l'IA un large éventail de parties prenantes, dont les pouvoirs publics, les entreprises privées et la société civile. Faisant écho à cette proposition, d'autres membres ont souligné la nécessité pour les pouvoirs publics de se montrer plus réactifs face aux évolutions et aux progrès rapides de cette technologie, notamment en instaurant un dialogue avec diverses parties prenantes de manière à en comprendre les conséquences sur différents segments de la société et différents secteurs. Un autre membre du G7 a également fait valoir l'impératif d'une harmonisation et d'une collaboration à l'échelle internationale afin de garantir la fiabilité de ces technologies ainsi qu'une approche commune quant à leur gouvernance. Cela est d'autant plus pertinent que les technologies d'IA sont mises au point puis déployées partout dans le monde, dans des États démocratiques aussi bien que dans des régimes antidémocratiques.

TYPES DE MESURES QUE LE G7 POURRAIT RECOMMANDER AUX POUVOIRS PUBLICS

Les participants considèrent que les mesures que le G7 pourrait recommander de mettre en œuvre sans tarder, parmi celles proposées dans le questionnaire, sont la mise à disposition d'outils efficaces permettant de garantir la sécurité, d'assurer un contrôle de la qualité et de contribuer au renforcement des capacités et de la confiance, ainsi que l'ouverture d'un dialogue (Graphique 2.8).

En ce qui concerne les mesures d'incitation (« Develop incentives... »), un membre du G7 a fait observer que dans de nombreux cas, les outils de base permettant d'évaluer ou de gérer les risques n'existent pas encore. Il importe donc au préalable de poursuivre les activités de R-D en vue d'améliorer l'analyse des risques.

GRAPHIQUE 2.8. TYPES DE MESURES DONT LE G7 POURRAIT RECOMMANDER LA MISE EN ŒUVRE SANS TARDER



Note : Le chiffre indiqué correspond à la somme des réponses données par les sept participants à la question : « Quel type de mesures le G7 pourrait-il recommander pour évaluer collectivement les progrès de l'IA générative et les moyens d'en tirer parti tout en gérant les risques induits par cette technologie ? ».

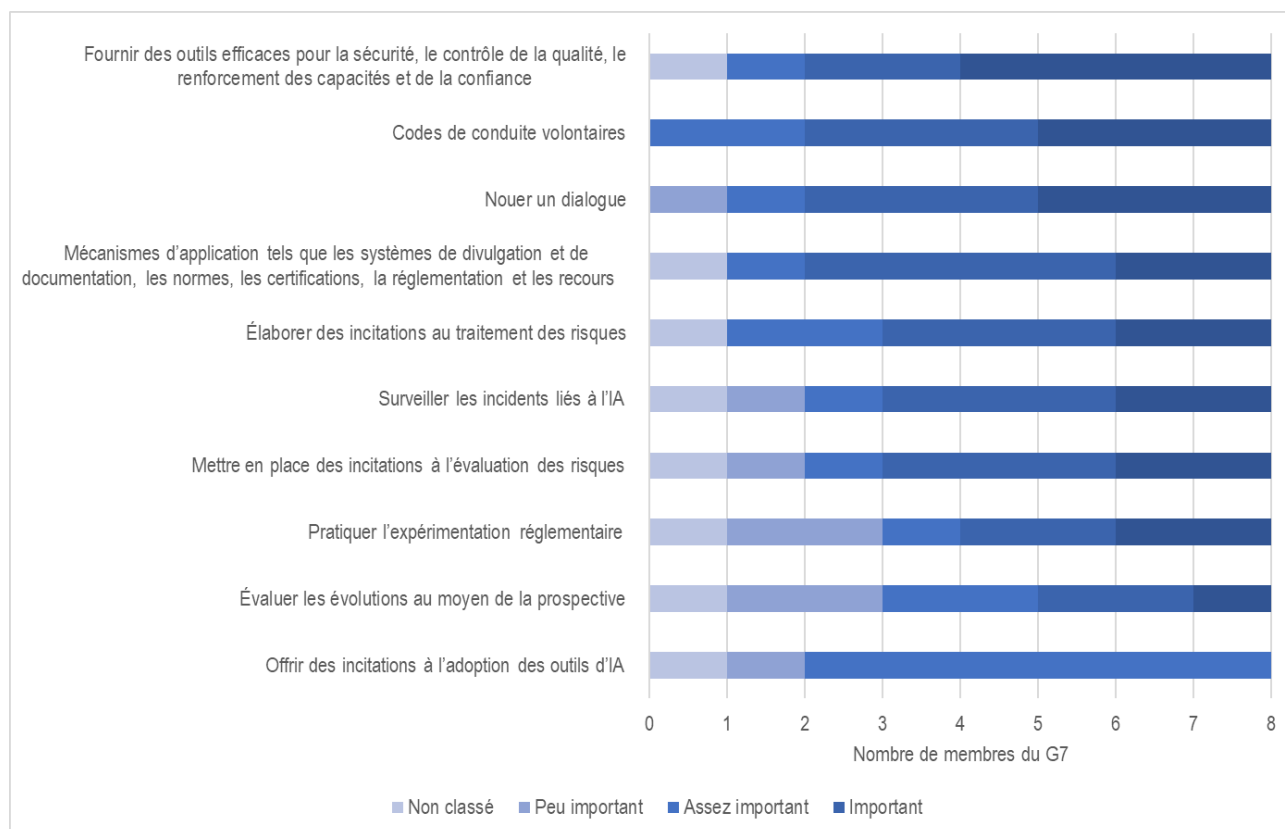
*Les participants considèrent que les mesures les plus importantes que le G7 pourrait recommander de mettre en œuvre, parmi celles proposées dans le questionnaire, sont la mise à disposition d'outils efficaces permettant de garantir la sécurité, d'assurer un contrôle qualité et de contribuer au renforcement des capacités et de la confiance, ainsi que la mise en place de codes de conduite non contraignants (Figure 2.9 *Error! Not a valid bookmark self-reference.*).*

Selon un membre du G7, la collaboration sur des projets concrets avec des experts en intelligence artificielle afin de tirer parti des possibilités qu'offre l'IA et de gérer les risques inhérents à cette technologie est une mesure importante qu'il conviendrait de mettre en œuvre sans tarder. Pour d'autres membres du G7, relèvent aussi de cette catégorie les actions de communication, le partage des connaissances et des bonnes pratiques, et l'ouverture d'un dialogue avec l'ensemble des parties prenantes concernées à l'échelle nationale comme internationale. Les participants ont en outre indiqué que l'adoption d'approches réglementaires était l'une des mesures les plus importantes que le G7 pourrait recommander pour évaluer collectivement les évolutions de l'IA générative et les risques posés par cette technologie.

Il a également été proposé de recenser les initiatives des pouvoirs publics dans le domaine de l'IA par l'intermédiaire de l'Observatoire OCDE des politiques relatives à l'IA en vue de contribuer à l'élaboration d'une base de données factuelles qui servira à nourrir les débats sur les politiques publiques. Bien que les initiatives en lien avec l'IA générative soient déjà consignées dans l'Observatoire OCDE des politiques relatives à l'IA par les pays, il est prévu de créer une catégorie et une section spéciales afin de les identifier plus facilement.

En ce qui concerne les outils permettant de garantir la sécurité, d'assurer un contrôle de la qualité, et de contribuer au renforcement des capacités et de la confiance, l'OCDE a publié en avril 2023 le [Catalogue d'outils et d'indicateurs à l'appui d'une IA digne de confiance](#). Le catalogue offre aux professionnels de l'IA du monde entier un cadre leur permettant de partager et de comparer des outils et de mettre à profit leurs travaux respectifs pour élaborer des bonnes pratiques mondiales et accélérer le processus de mise en œuvre des Principes de l'OCDE sur l'IA. Il peut s'agir d'outils de nature procédurale, technique ou éducative, classés en fonction de leur objectif (c'est-à-dire le Principe de l'OCDE sur l'IA auquel ils sont censés se rapporter). Le catalogue compte à l'heure actuelle 585 outils à l'appui d'une IA digne de confiance, dont [28 outils](#) qui concernent plus précisément l'IA générative.

GRAPHIQUE 2.9. MESURES LES PLUS IMPORTANTES QUE LE G7 POURRAIT RECOMMANDER



Note : Le chiffre indiqué correspond à la somme des réponses données par les huit participants à la question : « *Quel type de mesures le G7 pourrait-il recommander pour évaluer collectivement les progrès de l'IA générative et les moyens d'en tirer parti tout en gérant les risques induits par cette technologie ?* »

2.4. INITIATIVES NATIONALES ET RÉGIONALES DES MEMBRES DU G7

Les membres du G7 tirent parti des possibilités offertes par l'IA générative (Tableau 2.1) ou font face aux défis qu'elle induit (Tableau 2.2), dans le cadre soit d'initiatives existantes, soit de nouvelles mesures.

Les **États-Unis** ont répondu que nombre de leurs politiques publiques, décrets et approches en vigueur (par exemple, le cadre NIST de gestion des risques liés à l'IA) s'appliquaient aussi à l'IA générative. Plusieurs des engagements volontaires pris par les grandes entreprises d'IA portent notamment sur des questions propres à l'IA générative (celui de développer des systèmes de tatouage numérique en est un exemple).

Le **Canada** a indiqué que certains de ses ministères et organismes étudiaient actuellement la possibilité de lancer de nouvelles initiatives à l'appui de la recherche-développement dans le domaine de l'IA générative, et qu'il communiquerait de plus amples informations à ce sujet au fur et à mesure de l'avancement des travaux.

Les réglementations et politiques en vigueur dans l'**Union européenne**, notamment en matière de protection des données, de mésinformation, de pratiques commerciales déloyales, de protection du droit d'auteur et de services numériques, s'appliquent pleinement à l'IA générative. En outre, dans le cadre des négociations législatives menées actuellement en lien avec la législation de l'UE sur l'IA, des discussions sont en cours pour renforcer les dispositions portant spécifiquement sur l'IA générative.

TABLEAU 2.1. APERÇU DES INITIATIVES NATIONALES ET RÉGIONALES VISANT À TIRER PARTI DES POSSIBILITÉS OFFERTES PAR L'IA GÉNÉRATIVE

Membre	Initiative existante	Description/actions en matière d'IA générative	Type d'IA générative	Nouvelle initiative	Description	Type d'IA générative
Allemagne	Stratégie nationale en matière d'IA	Cadre pour une action publique globale sur le développement et l'application futurs de l'IA	Multimodale	Grands modèles européens d'AI (LEAM)	Initiative de la German AI Association visant à favoriser le développement des grands modèles de langage dans l'UE et en Allemagne	Texte
	FAIR Forward	Initiative de développement à l'appui d'une application plus ouverte et plus durable de l'IA dans les économies en développement et émergentes		F13	Grand modèle de langage basé sur le modèle Luminous utilisé par l'administration du land de Bade-Wurtemberg	Texte
	Lignes directrices pour la promotion d'une IA au service de l'intérêt commun	Renforcement de l'utilisation des données et des technologies faisant appel à l'IA au service de l'intérêt commun		BeKI	Centre consultatif sur l'IA dans le secteur public	
	Centres d'excellence pour la recherche sur l'IA	Six centres nationaux de compétences en matière de recherche sur l'IA, qui ont pour objectif de favoriser l'excellence dans ce domaine	Multimodale			
	Centre pour la qualité et l'innovation en matière d'IA	Mise au point d'approches fondées sur des tests pratiques de l'IA pour garantir l'existence de normes relatives aux applications et assurer la qualité de celles-ci	Multimodale			
	Civic Coding – KI für das Gemeinwohl	Structures favorisant l'émergence d'innovations sociales autour de l'IA				
	Plattform Lernende Systeme – Plateforme de l'Allemagne pour l'IA	Forum d'échange et de coopération	Multimodale			

Membre	Initiative existante	Description/actions en matière d'IA générative	Type d'IA générative	Nouvelle initiative	Description	Type d'IA générative
Canada	Stratégie pancanadienne en matière d'IA	Stratégie centrée sur des investissements ciblant les talents, la capacité de recherche, la commercialisation et la normalisation		Orientations sur l'utilisation de l'IA générative dans la fonction publique fédérale		
	Directive sur la prise de décisions automatisée	Directive soutenue par l'outil d'évaluation de l'incidence algorithmique, qui régit l'utilisation des systèmes d'IA au sein de la fonction publique fédérale		Proposition de code de pratique canadien pour l'IA générative	Organisation par le Canada de tables rondes afin de recueillir les observations des parties prenantes concernant une proposition de code de pratique canadien sur l'IA générative, qui fournira des orientations volontaires aux entreprises développant et utilisant des systèmes d'IA et les aidera à préparer leurs processus et leurs produits avant que la réglementation formelle prenne effet	
États-Unis				Engagements volontaires	Entre autres, un engagement des grandes entreprises spécialisées dans l'IA à mettre au point et à déployer des systèmes d'IA de pointe pour aider à relever les plus grands défis auxquels la société est confrontée	Multimodale
				Groupe de travail public sur l'IA générative du NIST	Définition d'un cadre NIST de gestion des risques liés à l'IA pour les systèmes d'IA générative	Multimodale
				Groupe de travail sur l'IA générative du Conseil consultatif du Président sur la science et la technologie (PCAST)	Élaboration de recommandations destinées au Président sur la marche à suivre pour que ces technologies soient développées et déployées de façon aussi équitable, responsable et sûre que possible	Multimodale

Membre	Initiative existante	Description/actions en matière d'IA générative	Type d'IA générative	Nouvelle initiative	Description	Type d'IA générative
France	Stratégie nationale en matière d'IA	Accès des startups au supercalcul	Multimodale	EDIC pour le NLP	Initiative européenne pour la création d'un consortium international pour le traitement automatique du langage naturel (actuellement en cours d'extension pour inclure l'IA générative multimodale)	Multimodale/texte
		Appel à projets pour la création de communs numériques pour l'IA générative en français	Multimodale			
		Grand défi international sur l'IA générative visant à évaluer l'état de l'art s'agissant des grands modèles de langage dans le monde	Texte			
Italie	Fonds de développement des technologies et applications faisant intervenir l'intelligence artificielle, la technologie des chaînes de blocs et l'internet des objets	Fonds destiné aux entreprises, y compris aux PME, qui contribue au financement de projets de recherche industrielle et d'innovation visant à développer des technologies et applications fondées sur l'IA, la technologie des chaînes de blocs et l'internet des objets	Multimodale	Révision du Plan stratégique en faveur de l'intelligence artificielle de 2021 (achèvement et mise à disposition pour consultation ouverte prévus pour fin septembre 2023)	Mise à jour de la stratégie nationale en matière d'IA au vu de la législation de l'UE sur l'IA et du développement de l'IA générative	Multimodale
	Programme national de doctorats en IA	Cinq cursus fédéraux de doctorat rassemblant 61 universités et établissements de recherche	Multimodale	Nouvelle gouvernance : Autorité nationale responsable des questions liées à l'IA au sein de l'Agence pour l'Italie numérique (AgID), placée sous la supervision du Département de la Transformation numérique	À l'avenir, coordination par l'Autorité de l'ensemble des nouvelles initiatives stratégiques, tout particulièrement dans le domaine de l'IA générative	
				Fonds de capital-risque d'entreprise pour que les startups spécialisées dans l'IA développent des solutions pour le secteur public	Mise en œuvre de la « Stratégie nationale de 2021 en matière d'IA », objectif n° 5 : « Développer des mesures et des services fondés sur l'IA dans le secteur public en favorisant l'innovation »	Multimodale

Membre	Initiative existante	Description/actions en matière d'IA générative	Type d'IA générative	Nouvelle initiative	Description	Type d'IA générative
				Assistant virtuel personnalisé – Institut national de sécurité sociale (INPS)	Assistant virtuel utilisant l'IA générative qui permet d'améliorer l'expérience de recherche sur le site web de l'INPS. Version bêta de cet assistant facilitant la soumission de questions spécifiques concernant une mesure de sécurité sociale réservée aux femmes appelée « Opzione Donna » (« Option Femme »)	Multimodale
				Tri et classification automatiques des courriels certifiés – INPS	Amélioration du flux de transmission des millions de courriels certifiés reçus chaque semaine	Texte
				Projet de recherche sur l'avenir de l'IA (FAIR)	Réseau de recherche visant à contribuer à la réflexion sur les thématiques, les méthodes, les modèles et les technologies, ainsi que les règles éthiques et juridiques, liés à la création de systèmes d'IA centrés sur l'humain, qui comprend quatre établissements de recherche et 14 universités	
Japon	Stratégie de 2022 en matière d'IA	Stratégie nationale en matière d'IA présentant un programme d'action complet en lien avec l'IA		Équipe stratégique sur l'IA et Conseil stratégique sur l'IA	Organismes publics et groupes d'experts examinant un large éventail de questions en lien avec l'IA générative	Multimodale
				Synthèse provisoire des problématiques liées à l'IA	Synthèse des problématiques liées à l'IA (principalement générative) par le Conseil de stratégie des technologies liées à l'IA	Multimodale
				Utilisation de l'IA	Accord sur l'utilisation de l'IA générative dans l'administration	Multimodale

Membre	Initiative existante	Description/actions en matière d'IA générative	Type d'IA générative	Nouvelle initiative	Description	Type d'IA générative
				Renforcement des capacités de développement en matière d'IA	Appui au renforcement des ressources de calcul pour les entreprises privées	Multimodale
Royaume-Uni	Livre blanc sur la réglementation de l'IA	Approche contextuelle, proportionnée et adaptable de la réglementation de l'IA, qui se fonde sur l'expertise des régulateurs existants	Multimodale	Groupe de réflexion sur les modèles de fondation	Pilotage, par le groupe de réflexion, de travaux de recherche en matière de sécurité de l'IA afin de faire progresser le développement sécurisé et fiable de modèles de fondation	Multimodale
	Portefeuille de techniques d'assurance de l'IA du Centre pour l'éthique des données et l'innovation (CDEI)	Études de cas sur les techniques d'assurance de l'IA appliquées par des organisations de différents secteurs	Multimodale	Sommet mondial du Royaume-Uni sur la sécurité de l'IA	Examen des risques que présente l'IA et discussions sur les moyens de les atténuer par une action coordonnée à l'échelle internationale	Multimodale
	AI Standards Hub	Outils et informations pratiques pour favoriser l'adoption et l'élaboration de normes en matière d'IA	Multimodale			
UE	Législation de l'UE sur l'IA	Obligations en matière de transparence, bacs à sable réglementaires	Multimodale	Installations d'essai et d'expérimentation (TEF)	Co-financement des TEF par la Commission européenne pour soutenir les développeurs de l'IA	Ouverture à toutes les technologies utilisant l'IA
	Plan coordonné dans le domaine de l'IA	Stratégie ciblant l'IA et les domaines d'action prioritaires élaborée en coopération avec les États membres, qui comprend des mesures visant à promouvoir l'accès à une plus grande puissance de calcul, la microélectronique, les TEF et les pôles d'innovation numérique		Consortium pour une infrastructure numérique européenne		
	Entreprise commune pour le calcul à haute performance européen (EuroHPC JU)					

Tableau 2.2. Aperçu des initiatives nationales et régionales visant à faire face aux risques que présente l'ia générative

Membre	Initiative	Description	Type d'IA générative
Canada	Loi sur l'intelligence artificielle et les données (LIAD)	Proposition relative à l'élaboration d'un cadre réglementaire national fondé sur les risques s'appliquant à la conception, au développement et à l'utilisation responsables de l'IA dans le secteur privé	Multimodale
	Proposition de code de pratique canadien pour l'IA générative	Organisation par le Canada de tables rondes afin de recueillir les observations des parties prenantes concernant une proposition de code de pratique canadien sur l'IA générative, qui fournira des orientations volontaires aux entreprises développant et utilisant des systèmes d'IA et les aidera à préparer leurs processus et leurs produits avant que la réglementation formelle prenne effet	Multimodale
	Guide du SCT sur l'utilisation de l'IA générative au sein du gouvernement du Canada	Formulation d'orientations relatives à l'utilisation des outils d'IA générative destinées aux institutions fédérales	Multimodale
États-Unis	Engagements volontaires	Engagements volontaires des plus grandes entreprises d'IA qui visent à favoriser la sûreté, la sécurité et la transparence au niveau du développement des technologies d'IA	Multimodale
	Groupe de travail public sur l'IA générative du NIST	Définition d'un cadre NIST de gestion des risques liés à l'IA pour les systèmes d'IA générative	Multimodale
	Groupe de travail sur l'IA générative du Conseil consultatif du Président sur la science et la technologie (PCAST)	Élaboration de recommandations destinées au Président sur la marche à suivre pour que ces technologies soient développées et déployées de façon aussi équitable, responsable et sûre que possible	Multimodale
Italie	Document d'orientation sur l'évaluation, l'audit et la gestion des risques	Document d'orientation sur l'évaluation, l'audit et la gestion des risques produit dans le cadre d'une collaboration de divers acteurs de la réglementation	Multimodale
Japon	Synthèse provisoire des problématiques liées à l'IA	Synthèse des problématiques liées à l'IA (principalement générative) par le Conseil de stratégie des technologies liées à l'IA	Multimodale
	Réponses à apporter face aux risques liés à l'IA	Travaux sur des lignes directrices uniformes pour les acteurs économiques, synthèse des problématiques liées aux droits de propriété intellectuelle, définition de lignes directrices pour le recours à l'IA générative dans le domaine éducatif, etc.	Multimodale
Royaume-Uni	Livre blanc sur la réglementation de l'IA	Approche contextuelle, proportionnée et adaptable de la réglementation de l'IA, qui se fonde sur l'expertise des régulateurs existants	Multimodale
	Portefeuille de techniques d'assurance de l'IA du Centre pour l'éthique des données et l'innovation (CDEI)	Études de cas sur les techniques d'assurance de l'IA appliquées par des organisations dans un éventail de secteurs	Multimodale
	Sommet mondial du Royaume-Uni sur la sécurité de l'IA	Examen des risques que présente l'IA et discussions sur les moyens de les atténuer par une action coordonnée à l'échelle internationale	Multimodale
	AI Standards Hub	Outils et informations pratiques pour favoriser l'adoption et l'élaboration de normes en matière d'IA	Multimodale
UE	Législation sur l'intelligence artificielle	Législation de l'UE (en cours d'examen)	Multimodale

INITIATIVES NATIONALES ET RÉGIONALES DANS LE DOMAINE DE L'IA GÉNÉRATIVE, PAR MEMBRE DU G7

Allemagne

- Stratégie nationale en matière d'IA : cadre pour une action publique holistique sur le développement et l'application futurs de l'IA.
- FAIR Forward : initiative liée au développement visant à favoriser une application plus ouverte et durable de l'IA, aux côtés des économies en développement et émergentes.

- Grands modèles européens d'AI (LEAM) : initiative de la *German AI Association* visant à favoriser le développement des LLM dans l'UE et en Allemagne.
- Centres d'excellence pour la recherche sur l'IA : le ministère fédéral de l'Éducation et de la Recherche a créé six centres nationaux de compétences en matière de recherche sur l'IA afin de favoriser l'excellence et la compétitivité et afin d'endosser un rôle de premier plan en matière de recherche sur l'IA.
- Codage civique — *KI für das Gemeinwohl* : réseau d'innovation visant à créer des structures favorisant l'émergence d'innovations sociales et une appropriation sociale plus large de l'IA.
- Ligne directrice pour favoriser l'utilisation d'une IA au service de l'intérêt commun : financement de travaux de recherche et de mise en œuvre et de projets de modèles afin de renforcer l'utilisation de données et de technologies faisant appel à l'IA au service de l'intérêt général.
- F13 : LLM basé sur le modèle Luminous. Il est utilisé par les pouvoirs publics du land de Bade-Wurtemberg.
- Systèmes d'apprentissage (*Lernende Systeme*) — plateforme de l'Allemagne pour l'IA : forum d'échange et de coopération rassemblant des acteurs des sciences, de l'industrie et de la société et visant à conforter l'Allemagne dans un positionnement de chef de file international dans le domaine des technologies.
- BeKI : centre consultatif sur l'IA dans le secteur public.

Canada

- Loi sur l'intelligence artificielle et les données (LIAD) : Déposée au Parlement canadien en juin 2022, la LIAD propose l'élaboration d'un cadre réglementaire national fondé sur les risques s'appliquant à la conception, au développement et à l'utilisation responsables de l'IA dans le secteur privé.
- Proposition de code de pratique canadien sur l'IA générative : Le Canada organise des tables rondes afin de recueillir les observations des parties prenantes concernant une proposition de code de pratique canadien sur l'IA générative. Ce code fournira des orientations volontaires aux entreprises qui développent et utilisent des systèmes d'IA, et les aidera à préparer leurs processus et leurs produits avant que la réglementation formelle prenne effet.
- Guide du SCT sur l'utilisation de l'IA générative au sein du gouvernement du Canada : Le guide du SCT se veut un document évolutif destiné à être affiné au fil du temps, qui est conçu pour procurer aux institutions fédérales des orientations relatives à l'utilisation des outils d'IA générative. Il donnerait une vue d'ensemble de l'IA générative, identifierait les défis et les préoccupations liés à son utilisation, présenterait des principes pour son utilisation responsable, et proposerait des considérations d'action et des bonnes pratiques en la matière.

États-Unis

- Engagements sur la base du volontariat : pour exploiter pleinement le potentiel de l'IA, les États-Unis encouragent le secteur de l'IA à respecter les plus hautes exigences pour garantir que l'innovation ne se fasse pas aux dépens des droits et de la sécurité des Américains. La Maison Blanche a obtenu des plus grandes entreprises d'IA, sur la base du volontariat, des engagements qui visent à favoriser la sûreté, la sécurité et la transparence au niveau du développement des technologies d'IA. Ces engagements mettent en relief trois principes fondamentaux pour l'avenir de l'IA — la sûreté, la sécurité et la confiance —, et ils marquent une étape cruciale dans le développement d'une IA responsable.

- Groupe de travail public sur l'IA générative de l'Institut national des normes et de la technologie (NIST) : ce groupe de travail s'appuie sur le succès du Cadre du NIST sur la gestion des risques liés à l'IA pour faire face à l'évolution rapide des technologies d'IA. Le Groupe de travail public sur l'IA générative permettra d'évoquer les possibilités et les défis liés aux outils d'IA capables de générer du contenu, qu'il s'agisse de code, de textes, d'images, de vidéos ou de musique. Ce Groupe de travail aidera aussi le NIST à formuler de grandes orientations pour aider les organisations à faire face aux risques spécifiques découlant des technologies d'IA générative.
- Groupe de travail du PCAST sur l'IA générative : le Conseil consultatif du Président sur la science et la technologie (PCAST) s'est doté d'un Groupe de travail sur l'IA générative qui contribuera à examiner les grandes possibilités et les grands risques de ce secteur et à formuler des orientations sur la marche à suivre pour que ces technologies soient développées et déployées de façon aussi équitable, responsable et sûre que possible. Le Groupe de travail du PCAST sur l'IA générative a pour ambition de prolonger l'action déjà menée en la matière en répertoriant d'autres besoins et possibilités et en formulant, à l'intention du Président américain, des recommandations sur l'action à mener en conséquence.

France

- Stratégie nationale pour l'intelligence artificielle : cette stratégie consiste à accompagner le lancement d'initiatives en matière d'IA générative en France en facilitant l'accès des startups à de la puissance de calcul, en lançant un appel à projets en vue de la création de communs numériques pour l'IA générative en français (bases de données, modèles applicatifs utilisant l'IA, etc.) et en dévoilant un défi mondial en matière d'IA générative visant à évaluer l'état de l'art sur le plan des grands modèles de langage (LLM) du monde entier.
- EDIC pour le NLP : initiative européenne visant à créer un consortium international pour le traitement du langage naturel (en cours d'extension pour inclure l'IA générative multimodale). Cette initiative permettra de mettre en place un ensemble européen de bases de données, de mutualiser la puissance de calcul européenne pour les projets prometteurs, d'offrir un incubateur à des startups, de financer des projets de recherche, etc.

Italie

- Mise en œuvre de la Stratégie nationale 2021 en matière d'IA, objectif n° 5 : « Élaborer des politiques et des services fondés sur l'IA dans le secteur public en stimulant l'innovation dans le secteur public ».
- Création d'un fonds de capital-risque pour que les startups spécialisées dans l'IA développent des solutions destinées au secteur public : ce fonds (qui commencera à fonctionner en 2024) investira dans des startups susceptibles de développer des technologies de rupture pour l'automatisation de processus au sein des institutions publiques. L'investissement s'élève à 600 millions EUR. Le fonds englobe un mécanisme institutionnel qui vise à favoriser le dialogue entre les startups et les administrations publiques afin de mieux faire comprendre les besoins du secteur public. Le fonds est placé sous la responsabilité de l'Agence de la cybersécurité nationale (ACN) et du Département de la Transformation numérique.
- Création d'une autorité nationale responsable des questions liées à l'IA au sein de l'Agence pour l'Italie numérique (AgID), sous la supervision du Département de la Transformation numérique. L'Autorité coordonnera l'ensemble des nouvelles initiatives stratégiques, tout particulièrement dans le domaine de l'IA générative.
- Fonds de développement des technologies et applications faisant intervenir l'intelligence artificielle, les chaînes de blocs et l'internet des objets : fonds consacré aux entreprises, y compris les PME. Il contribue au financement de projets de recherche industrielle et d'innovation visant à

développer des applications fondées sur l'intelligence artificielle, les chaînes de blocs et l'internet des objets.

- L'assistant virtuel personnalisé de l'Institut national de la sécurité sociale (INPS) aide les usagers à s'orienter parmi les nombreux services proposés sur le portail de l'INPS.
- Tri et classification automatiques des courriels certifiés envoyés à l'Institut national de la sécurité sociale (INPS) : ce projet vise à améliorer le flux de transmission des millions de courriels certifiés envoyés à l'INPS, grâce à un système permettant d'analyser automatiquement le contenu du courriel reçu et de le faire parvenir à l'agent chargé de fournir les réponses correspondantes.
- Programme national italien de doctorats en intelligence artificielle : le programme national italien de doctorats en intelligence artificielle consiste en cinq grands cursus de doctorat qui rassemblent 61 universités et institutions de recherche. Les cinq cursus de doctorat comportent un socle commun d'enseignement consacré à l'étude des fondements et des évolutions de l'IA, et chacun propose, en plus, un volet spécialisé consacré à un secteur stratégique de recours à l'IA. Chaque cursus de doctorat est organisé par une université chef de file, en collaboration avec le Conseil national de la recherche (CNR).
- Projet de recherche sur l'avenir de l'intelligence artificielle (FAIR) : le projet FAIR vise à contribuer à la réflexion sur les thématiques, les méthodes, les modèles, les technologies et même les règles éthiques et juridiques liés à la création de systèmes d'IA centrés sur l'humain. Le réseau de recherche FAIR s'étend sur l'ensemble du pays et rassemble quatre institutions de recherche (CNR, Fondazione Bruno Kessler, INFN et IIT), 14 universités (Politecnico di Milano, Politecnico di Torino, Sapienza, Scuola Normale Superiore, SISSA, Università Bocconi, Università Campus Biomedico di Roma, Università della Calabria, Università di Bari, Università di Bologna, Università di Catania, Università di Napoli "Federico II", Università di Pisa et Università di Trento) et sept entreprises (Bracco, Deloitte, Expert. ai, Intesa Sanpaolo, Leonardo, Lutech et STMicroelectronics).
- Document d'orientation sur l'évaluation, l'audit et la gestion des risques : plusieurs acteurs du domaine de la réglementation dialoguent afin de produire un document d'orientation sur l'évaluation, l'audit et la gestion des risques, l'objectif étant de mieux comprendre les positions respectives des startups, des syndicats professionnels et du tiers-secteur.

Japon

- Stratégie 2022 en matière d'IA : lancée en avril 2022 et fondée sur les principes de la dignité, de la diversité et de la durabilité, cette stratégie vise à contribuer à la résolution des problématiques mondiales grâce à l'avènement de la Société 5.0, ainsi qu'à présenter un programme d'action complet en lien avec l'IA pour aider le Japon à relever ses propres défis sociaux et à renforcer la compétitivité de son industrie.
- Équipe stratégique sur l'IA et Conseil stratégique sur l'IA : en avril 2023, une équipe stratégique sur l'IA composée de professionnels des ministères et institutions concernés a été créée afin d'examiner la marche à suivre pour traiter un large éventail de problématiques liées à l'IA générative. De plus, en mai 2023, un Conseil stratégique sur l'IA composé d'experts a été créé afin d'apporter des éclairages issus d'horizons divers non seulement sur les aspects technologiques, mais aussi sur les aspects juridiques et éthiques liés à l'IA.
- Synthèse provisoire des problématiques liées à l'IA : au vu de l'accélération récente des évolutions technologiques et dans le prolongement du Sommet d'Hiroshima du G7, le Conseil stratégique sur l'IA a rédigé, fin mai 2023, une synthèse des questions liées à l'IA (principalement générative). Le Japon a lancé des discussions sur les réponses à apporter face aux risques liés à l'IA, sur le recours à l'IA et sur le renforcement des capacités de développement en matière d'IA. Il anime

également des discussions internationales, y compris dans le cadre du processus d'Hiroshima sur l'IA, dans le contexte de la présidence japonaise du G7.

- Réponses à apporter face aux risques liés à l'IA : travaux sur des lignes directrices uniformes pour les acteurs économiques, synthèse des problématiques liées aux droits de propriété intellectuelle, définition de lignes directrices pour le recours à l'IA générative dans le domaine éducatif, etc.
- Recours à l'IA : accord sur le recours à l'IA générative dans le secteur public.
- Renforcement des capacités de développement en matière d'IA : appui au renforcement des ressources de calcul pour les entreprises privées.

Royaume-Uni

- Livre blanc sur la réglementation de l'IA : ce Livre blanc présente, pour le Royaume-Uni, une approche contextuelle, proportionnée et adaptable de la réglementation de l'IA. Il se fonde sur l'expertise des régulateurs existants, qu'il encourage à examiner la marche à suivre pour encadrer au mieux l'IA dans leurs secteurs respectifs. Ce Livre blanc permettra au Royaume-Uni de trouver le bon équilibre entre la prise en compte des risques et la pleine mise à profit des possibilités offertes par l'IA.
- Portefeuille de techniques d'assurance de l'IA du Centre pour l'éthique des données et l'innovation (CDEI) : ce portefeuille présente des études de cas de techniques d'assurance de l'IA appliquées par des organisations utilisatrices de technologies de pointe dans un éventail de secteurs. Il s'agit d'une ressource précieuse, qui aidera ceux qui développent et qui achètent des systèmes d'IA à comprendre comment des techniques d'assurance de l'IA peuvent les aider à mesurer, évaluer et faire connaître la fiabilité des systèmes d'IA, et à comprendre comment les techniques s'articulent avec les principes d'encadrement énoncés dans le Livre blanc du Royaume-Uni sur la réglementation de l'IA.
- Groupe de réflexion sur les modèles de fondation : ce groupe de réflexion dirigera des recherches cruciales en matière de sécurité de l'IA afin de faire progresser le développement sécurisé et fiable de modèles de fondation tout en tirant parti des possibilités extraordinaires que ces derniers ouvrent. Ce groupe de réflexion bénéficie d'un financement public initial de 100 millions GBP.
- Sommet mondial du Royaume-Uni sur la sécurité de l'IA : ce sommet sera l'occasion d'examiner les risques que présente l'IA, y compris les systèmes de pointe, et de réfléchir aux moyens de les atténuer par une action coordonnée à l'échelle internationale. Il constituera aussi pour les pays un espace de collaboration pour poursuivre les travaux visant à définir une démarche commune d'atténuation des risques.
- AI Standards Hub : cette plateforme vise à favoriser l'adoption et l'élaboration de normes en matière d'AI en mettant à la disposition des entreprises, des régulateurs et des organisations de la société civile britanniques des outils et des renseignements concrets. Il faut, de plus, que ces acteurs appliquent avec efficacité les normes relatives à l'IA, et qu'ils contribuent à leur élaboration. La plateforme *AI Standards Hub* s'inscrit dans le cadre de la Stratégie nationale du Royaume-Uni en matière d'IA, et elle a pour objectif ultime d'accroître la contribution britannique à l'élaboration de normes techniques mondiales dans ce domaine.

UE

- Installations d'essai et d'expérimentation (TEF) : dans le cadre d'une coopération avec les États membres de l'UE, la Commission européenne cofinance des TEF afin d'aider les développeurs d'IA à mettre sur le marché avec plus d'efficacité une IA digne de confiance et afin de faciliter sa diffusion en Europe.

- Projet de règlement européen sur l'intelligence artificielle : consciente que les systèmes d'IA en général, et d'IA générative en particulier, risquent de remettre en cause un large éventail de domaines et de pratiques, la Commission européenne a proposé ce projet de règlement en 2021. Il s'agit de poser un cadre juridique visant à garantir que les citoyens puissent faire confiance aux produits issus de l'IA.

RÉFÉRENCES

- OCDE (2023), *AI language models: Technological, socio-economic and policy considerations*, Éditions OCDE, <https://doi.org/10.1787/13d38f92-en>. [1]
- OCDE (À paraître), *Generative Artificial Intelligence. Select policy considerations, Part 1*. [2]