Recommandations de FLI pour le Sommet pour l'Action sur l'IA

France, 10-11 février 2025

Imane Bello (Ima)
Responsable du Sommet pour l'action sur l'IA
ima@futureoflife.org

future of life INSTITUTE

Contentu

L'Institut pour le Futur de la Vie (FLI) a pour objectif de promouvoir les avantages de la technologie et de réduire les risques associés. FLI est l'un des principaux porte-paroles mondiaux en matière de réglementation de l'intelligence artificielle (IA) et a créé l'un des ensembles de principes les plus anciens et les plus influents, les principes d'Asilomar. FLI a tissé un vaste réseau composé des principaux chercheurs en IA au monde, issus du monde universitaire, de la société civile et du secteur privé.

Consulter ce document en ligne : futureoflife.org/fr/document/sommet-2025

Image de couverture : Palais de l'Élysée vu depuis les jardins, Paris. Journées européennes du patrimoine 2014. Wikimedia Commons

Publié le 6 février 2025 | Par l'Institut pour le Futur de la Vie (FLI)

Introduction

Monsieur le Président de la République,

Madame Anne Bouverot, Envoyée spéciale,

Nous tenons à vous remercier d'avoir pris l'initiative d'organiser le troisième sommet mondial sur l'IA et d'avoir clairement mis l'accent sur l'action. S'appuyant sur les progrès accomplis à Bletchley et à Séoul, ce sommet contribuera à ce que le développement et le déploiement de l'IA profitent à tous. À la suite des recommandations intermédiaires que nous avons transmises en novembre dernier, et à la lumière des récents développements en IA et de l'émergence de nouveaux modèles, vous trouverez ci-dessous nos recommandations finales pour le Sommet pour l'Action sur l'IA.

L'Institut pour le Futur de la Vie (FLI) est une organisation indépendante à but non lucratif dont l'objectif est de faire évoluer les technologies innovantes au profit du bien-être et de réduire les risques majeurs à grande échelle. En 2017, FLI a organisé une conférence à Asilomar, en Californie, dans le but de créer l'un des premiers outils de réglementation de l'intelligence artificielle (IA) : les « Principes d'Asilomar sur l'IA ».[i] Depuis, l'organisation est devenue l'un des principaux porte-parole en matière de réglementation de l'IA à Washington D.C. et à Bruxelles, et la principale voix de la société civile pour l'IA dans le cadre de la feuille de route du Secrétaire général des Nations unies pour la coopération numérique. Notre équipe participe également activement aux groupes d'experts de l'OCDE.

Lors des précédents sommets sur l'IA, FLI a contribué à fournir aux organisateurs des informations essentielles sur les dernières avancées en matière d'IA, ainsi que des recommandations sur la manière de renforcer la sécurité de l'IA. FLI entend maintenir le même niveau d'engagement lors du sommet qui se tiendra en France. Parmi nos travaux les plus récents figurent des recherches portant, par exemple, sur l'interprétabilité,[ii] les préoccupations et les mesures d'atténuation socio-techniques[iii] et les mesures efficaces d'atténuation des risques des systèmes d'IA à usage général,[iv] ainsi que l'octroi de subventions à des universitaires travaillant sur la réduction de la concentration du pouvoir dans l'industrie de l'IA.[v] Nous sensibilisons également le grand public par le biais de nos lettres d'information,[vi] en organisant chaque mois des petits-déjeuners sur

i Future of Life Institute. (2017, 8 11). Asilomar AI Principles. Récupéré 11 25, 2024, depuis https://futureoflife.org/open-letter/ai-principles/

ii Baek, D., Li, Y., & Tegmark, M. (2024). Generalization from Starvation: Hints of Universality in LLM Knowledge Graph Learning. 10.48550/ARXIV.2410.08255.

iii Aguirre, A., Dempsey, G., Surden, H., & Reiner, P. (2020). AI loyalty: A New Paradigm for Aligning Stakeholder Interests. SSRN Electronic Journal. 10.2139/ssrn.3560653.

iv Uuk, R., Brouwer, A., Dreksler, N., Pulignano, V., & Bommasani, R. (2024). Effective Mitigations for Systemic Risks depuis General-Purpose AI. SSRN Preprint. SSRN. Récupéré 11 25, 2024.

v Future of Life Institute. (n.d.). How to mitigate AI-driven power concentration. Récupéré 11 25, 2024, depuis https://futureoflife.org/grant-program/mitigate-ai-driven-power-concentration/

vi Bello, I. (n.d.). AI Action Summit Newsletter | Ima Bello | Substack. Récupéré 11 25, 2024, depuis https://aiactionsummit.substack.com/

la sécurité de l'IA à Paris[vii] et d'autres évènements, tels que le récent symposium sur la sécurité de l'IA à l'Université de la Sorbonne.[viii]

Nous pensons que le sommet devrait poursuivre trois objectifs principaux, qui s'alignent sur les trois catégories précédemment définies par le Président de la République[ix] :

- **Science.** Le sommet devrait contribuer à faciliter l'échange d'informations afin de parvenir à une compréhension commune de la gravité du risque mondial lié à l'IA, principalement le risque de perte de contrôle, et à élaborer des outils permettant de réduire ce risque ;
- **Solutions.** Le sommet devrait définir un cadre permettant d'assurer une réglementation internationale efficace en matière d'IA, menée par le secteur public et en complément des initiatives existantes, tout en maintenant le caractère spécifique du sommet en tant que forum permettant à tous les grands gouvernements (y compris la Chine) de discuter des questions de sécurité de l'IA à l'échelle internationale ;
- **Standards.** Le sommet devrait déboucher sur un engagement politique ferme en faveur de la mise en œuvre de nouveaux standards de sécurité en matière d'IA à l'échelle mondiale, notamment en ce qui concerne les systèmes d'IA à usage général de plus en plus puissants. Pour ce faire, la France devrait s'appuyer sur son expérience unique en matière de définition de standards internationaux, qui remonte au système international d'unités.[x]

En définitive, nous souhaitons que le sommet permette le développement d'une structure internationale souple et efficace pour la réglementation de l'IA, ouvrant ainsi la voie à une ère d'innovation et de prospérité sans précédent. Nous vous souhaitons tout le succès possible dans les préparatifs du sommet et nous nous tenons prêts à mettre à disposition toute notre expertise pour soutenir la mise en place d'une réglementation mondiale efficace en matière d'intelligence artificielle.

Sincèrement,

Professeur Anthony Aguirre
Directeur exécutif

Professeur Max Tegmark
Président

vii Future of Life Institute. (n.d.). Petits-déjeuners sur la sécurité de l'IA. Récupéré 11 25, 2024, depuis https://futureoflife.org/fr/petits-dejeuners-sur-la-securite-a-l-ai/

viii Sorbonne Université. (n.d.). AI Safety Symposium. Récupéré 11 25, 2024, depuis https://sites.google.com/view/paiss2024/home

ix Élysée. (2024, 5 22). Rassemblement des plus grands talents français de l'IA. Récupéré 11 25, 2024, depuis https://www.elysee.fr/emmanuel-macron/2024/05/22/rassemblement-des-plus-grands-talents-francais-de-lia

x Various authors. (n.d.). Système décimal. Wikipédia. Récupéré 11 24, 2024, depuis https://fr.wikipedia.org/wiki/Syst%C3%A8me_d%C3%A9cimal

Analyse des progrès réalisés depuis le sommet de Bletchley

FLI a été invité à participer au sommet inaugural sur la sécurité de l'IA,[1] qui s'est tenu à Bletchley Park en novembre 2023, en tant qu'acteur clé représentant la société civile. Lors du sommet, nous avons présenté un certain nombre de recommandations, notamment la nécessité de parvenir à une compréhension commune de la gravité et de l'urgence des risques liés à l'IA, d'expliciter la nature mondiale du défi que représente l'IA et d'insister sur la nécessité d'une réponse mondiale unifiée.

À la suite du sommet de Bletchley Park, un programme d'essais de sécurité des modèles d'IA a été mis en place, dans le cadre duquel le gouvernement britannique participe aux tests des modèles avant et après leur déploiement.[2] Ils testent explicitement les moyens par lesquels l'IA avancée peut porter atteinte à la sécurité et à la sûreté nationales, ou causer des dommages à la société.

Depuis, des progrès considérables ont été accomplis en matière de réglementation de l'IA, tant au niveau national qu'international.

Lors du sommet de Séoul, qui s'est tenu en mai 2024, de nombreux pays ont appelé à approfondir les travaux scientifiques fondés sur des données concrètes concernant la sécurité de l'IA. Le rapport scientifique international sur la sécurité de l'IA avancée[3], dans sa version provisoire, a été publié et présenté lors du sommet. Il porte sur les risques malveillants (armes biologiques, deepfakes, escroqueries), les dysfonctionnements (failles dans la sécurité des produits, partialité, perte de contrôle) et les risques systémiques. Les engagements en matière de sécurité des modèles dits de frontière[4] ont également été établis et ont été acceptés par 16 entreprises du Moyen-Orient, d'Asie, d'Europe et du continent américain. Lors du sommet de Séoul, le ministère américain du commerce a publié son plan de collaboration mondiale entre certains instituts de sécurité et d'évaluation de l'IA.[5]

En août 2024, le règlement pionnier de l'UE en matière d'intelligence artificielle[6] est entré en vigueur. La réglementation des systèmes d'intelligence artificielle à usage général y a été incluse, en grande partie grâce à l'action de FLI[7] et d'autres acteurs de la société civile. L'entrée en vigueur de ce règlement a également permis de lancer le processus d'élaboration d'un code de bonnes pratiques concernant les systèmes d'IA à usage général les plus avancés, qui est actuellement bien engagé.

En septembre 2024, l'organe consultatif des Nations unies sur l'IA a publié son rapport final intitulé « Gouverner l'IA au bénéfice de l'humanité »,[8] qui a été immédiatement suivi par l'adoption historique du Pacte numérique mondial lors du Sommet de l'avenir. En novembre 2024, la première assemblée[9] de certains instituts de sécurité en matière d'IA s'est tenue à San Francisco. Lors de cette réunion, les membres du réseau se sont mis d'accord sur les quatre principes clés de leur mission,[10] la recherche, l'expérimentation, l'orientation et l'information, ainsi que le partage d'outils. Cette rencontre révèle la tendance croissante des pays à créer des organismes publics chargés de superviser le développement de l'IA et à commencer à se coordonner sur le plan technique.

Enfin, en janvier 2025, le Rapport international sur la sécurité de l'IA[11] a été publié. Soutenue par

30 pays, l'ONU et l'UE, sa publication illustre la coordination internationale croissante visant à mieux comprendre les risques associés aux systèmes d'IA à usage général, englobant les risques d'utilisation malveillante, érosion de l'intégrité de l'information, de perturbations majeures sur le marché du travail, les préoccupations environnementales et les risques de perte de contrôle.

Recommandations pour le programme du Sommet

Le gouvernement français a défini cinq pistes[12] pour le sommet pour l'action sur l'IA, qui témoignent clairement du haut niveau des ambitions de l'événement. Nous souhaitons que le programme final renforce la compréhension commune de la gravité des risques mondiaux liés à l'IA et confirme le rôle prépondérant du secteur public dans l'élaboration de la réglementation en matière d'IA (par opposition aux entreprises du secteur privé).

Science

Pour parvenir à une compréhension scientifique commune des risques liés à l'IA, FLI recommande d'ajouter trois éléments spécifiques au programme. Premièrement, les efforts actuels visant à développer une IA autonome polyvalente (des systèmes capables d'agir, de planifier et de poursuivre des objectifs) pourraient conduire à une perte de contrôle.[Voir 3] Compte tenu du risque critique que représente le développement de l'IA, le programme gagnerait à ce que des experts indépendants discutent de la probabilité de ce risque et des mesures d'atténuation potentielles.

Deuxièmement, le gouvernement français pourrait inclure une démonstration en direct de certaines fonctionnalités dangereuses de l'IA (telles que les capacités de persuasion dans un contexte électoral). Une telle démonstration en direct permettrait aux participants d'acquérir le même niveau de compréhension et constituerait une bonne introduction au Rapport scientifique international sur la sécurité de l'IA avancée,[Voir 3] et peut s'inspirer de la démonstration réalisée à San Francisco en novembre 2024.[13]

Troisièmement, il peut s'avérer intéressant de convier deux instituts de sécurité de l'IA de pays différents, tels que les instituts japonais et britannique de sécurité de l'IA, à évaluer conjointement un modèle d'IA avancé face à un risque donné. Lorsque les résultats de cette évaluation conjointe seront présentés lors du sommet, les participants pourront tirer des enseignements des similitudes et des différences entre les approches des deux administrations.

Solutions

Le Président de la République a appelé à la création de « solutions partagées pour relever les défis posés par l'IA ». Le monde ne pourra vraiment parvenir à ces solutions partagées qu'avec la participation de la Chine et des États-Unis. Si la recherche américaine fait l'objet d'une large compréhension internationale, il n'en va pas de même pour la recherche chinoise. Compte tenu du potentiel important et en pleine progression de la Chine - comme l'illustrent les modèles publiés au cours du mois dernier - et de la nature mondiale des risques liés à l'IA, un manque d'engagement pourrait conduire à l'émergence de systèmes d'IA dangereux qui affecteraient le monde entier.

Le sommet devrait donc permettre de renforcer les échanges avec la Chine,[14] notamment en lui donnant la possibilité, lors de la session ministérielle, de partager les meilleures pratiques et les domaines de recherche tels que ceux évoqués lors de la conférence mondiale sur l'IA qui s'est tenue à Shanghai en juillet 2024[15] et dans le cadre de la résolution de la troisième séance plénière.[16]

Standards

En tant que premier pays à accueillir un sommet sur l'IA après l'adoption du règlement européen sur l'IA, la France dispose d'une occasion unique de contribuer à transformer ce règlement européen en standard international pour la réglementation de l'IA. En vertu du règlement sur l'IA, la Commission européenne est tenue d'élaborer un « Code de bonnes pratiques » pour les systèmes d'IA à usage général. La Commission a déjà cherché à inscrire les échanges européens dans la conversation internationale en nommant l'auteur principal du rapport scientifique, Yoshua Bengio, à la direction de la rédaction de ce code de bonnes pratiques. Le sommet offre une nouvelle occasion d'uniformiser les règles en 1) veillant à ce que les engagements volontaires de Séoul soient intégrés dans le code de conduite de l'UE, 2) en encourageant la participation et le soutien du secteur privé dans la mise en œuvre du code, 3) en obtenant l'adhésion des pays non membres de l'UE pour que les engagements volontaires aient également une base législative.

Au-delà du code de bonnes pratiques, le sommet et la communauté internationale en général devraient codifier le consensus international émergent en établissant des lignes rouges pour la reproduction ou le développement autonome, la volonté de pouvoir et les manœuvres frauduleuses (c'est-à-dire les facteurs contribuant au risque de perte de contrôle).

Recommandations pour le Sommet

Recommandations pour l'ensemble des gouvernements participants

- Réévaluer et mettre à jour les stratégies nationales en matière d'IA afin de préparer la société à l'arrivée imminente de l'IA agentique, par exemple en investissant davantage dans le renforcement des infrastructures critiques et des institutions publiques clés contre les cyberattaques automatisées.
- Faire appel à des experts indépendants en gestion des risques pour l'élaboration de la politique en matière d'IA. Bien que l'IA constitue un défi inédit en matière de politique publique, de nombreux enseignements peuvent être tirés d'autres secteurs à haut risque, tels que l'aviation ou la biotechnologie.
- Établir des standards clairs permettant aux instituts nationaux chargés de la sécurité de l'IA ("AISI", "National AI Safety Institute") ou à des organismes similaires d'auditer les entreprises opérant dans leur juridiction.
- Mettre en place des dispositifs solides d'atténuation des risques, assortis de mesures incitatives claires. Faute de mesures gouvernementales proactives, le manque de confiance dans les systèmes d'IA risque de retarder l'adoption généralisée de l'IA.

- Définir clairement la mission des instituts de sécurité ou des structures similaires en fonction du contexte national, par exemple en mettant l'accent sur les avancées dans le domaine de la métrologie, sur l'utilisation industrielle de robots hautement performants ou sur l'utilisation d'ensembles de données nationaux uniques.
- Établir une cadence régulière de sommets pour discuter de la gouvernance internationale des systèmes d'IA à usage général et de leurs implications pour la sécurité mondiale, au vu des cycles de développement de plus en plus courts.
- Envisager, pour les prochains Sommets, une structure à deux volets : un premier consacré à la gouvernance des systèmes d'IA à usage général, impliquant principalement les pays disposant de capacités critiques en IA ainsi que les grandes entreprises opérant dans leur juridiction ; un second axé sur des opportunités d'intérêt public plus large.[17]

Recommandations pour les hôtes français

- Établir un équilibre entre les points de vue des parties prenantes en impliquant le secteur privé, les pouvoirs publics, les universitaires et la société civile, afin de parvenir à des actions significatives, atténuer les conflits d'intérêts et éviter le "blanchiment de sécurité" ("safety washing" en anglais).
- Demander aux entreprises signataires des engagements de Séoul en matière de sécurité de présenter leur cadre de sécurité suffisamment tôt avant le sommet et de rendre ces documents accessibles au public pour permettre un examen approfondi.
- Conformément aux conclusions de la consultation d'experts,[18] persuader les grandes entreprises de s'engager à faire évaluer les systèmes d'IA par des tiers indépendants.
- Veiller à l'intégration des engagements volontaires et des seuils de l'OCDE dans les futurs codes de bonnes pratiques de l'UE, afin d'éviter une fragmentation à l'échelle mondiale.
- Prévoir un espace lors de la session ministérielle pour que les parties prenantes chinoises puissent partager les meilleures pratiques et les domaines de recherche tels que ceux qui ont été abordés lors de la conférence mondiale sur l'IA qui s'est tenue à Shanghai en juillet 2024[Voir 15] et dans le cadre de la troisième séance plénière.[Voir 16]
- Désigner le(s) hôte(s) des prochains sommets afin d'en assurer la réussite et de garantir un transfert de connaissances harmonieux. Parmi les candidats, Singapour se révèle être un candidat de choix pour accueillir un futur sommet sur l'IA, en raison de sa relative neutralité diplomatique.
- Établir des lignes rouges en codifiant le consensus international émergent entre les principaux experts techniques et de réglementation de l'IA de Chine et d'Occident autour des cinq risques mondiaux critiques suivants : la reproduction ou le développement autonome, la volonté de puissance, l'aide au développement d'armes, les cyber-attaques et les pratiques frauduleuses.
- Renforcer le rôle de chef de file mondial de la France en mettant en place une fondation indépendante pour l'IA, associée à un fonds permettant la création de modèles à petite échelle, sur la base de données de qualité et pour des applications prédéfinies et spécifiques en rapport avec une IA d'intérêt public.

Recommandations pour la République populaire de Chine

- Informer les autres gouvernements des intentions de la Chine de mettre en place des systèmes de contrôle de la sécurité de l'IA (notamment par le biais de la troisième séance plénière[Voir 16]).
- Partager les enseignements tirés du cadre de réglementation de la sécurité de l'IA du TC260[19] et des réglementations chinoises en matière d'IA,[20] y compris le registre des algorithmes et les évaluations de sécurité et de sûreté.
- Faire intervenir les groupes chinois d'évaluation et de sécurité en matière d'IA et partager l'expertise d'éminents chercheurs chinois.
- Prendre part à un dialogue ouvert avec tous les pays participants, y compris les États-Unis. Les menaces mondiales liées à l'IA avancée, tout comme le changement climatique, nécessitent une coopération qui transcende la concurrence géopolitique.

Recommandations pour le Bureau européen de l'IA

- Fournir aux autres gouvernements des informations sur les futurs codes de bonnes pratiques de l'UE et sur toute autre information importante apparaissant au cours du processus de rédaction, en mettant l'accent sur le régime applicable aux systèmes d'IA à usage général.
- Aligner la législation de l'UE en matière d'IA et son code de bonnes pratiques sur les engagements volontaires de Séoul afin d'éviter la fragmentation de la réglementation mondiale en matière d'IA.
- Adapter le code de bonnes pratiques en fonction des dernières données scientifiques présentées lors du sommet.

Recommandations pour les États-Unis

- Suivre de près le respect des engagements volontaires par les grandes entreprises du secteur de l'IA et exercer les pressions nécessaires pour garantir le respect de ces engagements.
- Déterminer le risque que des systèmes produits aux États-Unis puissent être utilisés pour favoriser l'instabilité mondiale.
- Engager un dialogue ouvert avec tous les pays, y compris la Chine. Les menaces mondiales liées à l'IA avancée, tout comme le changement climatique, nécessitent une coopération qui transcende la concurrence géopolitique.

Recommandations pour les pays dotés d'infrastructures technologiques émergentes

- Exploiter stratégiquement les ressources linguistiques et culturelles nationales pour les ensembles de données de formation et d'évaluation de la sécurité de l'IA,[21] en assurant la création de valeur locale et la représentation culturelle, tout en veillant à ce que ces ressources restent la propriété des communautés concernées et soient protégées par un code source fermé afin d'éviter tout risque de compromission de la sécurité et de commercialisation non autorisée.
- Investir dans les capacités et l'expertise locales en matière de recherche sur la sécurité de l'IA, y compris les talents indigènes en matière de sécurité de l'IA, et négocier des accords de compensation équitables et transparents avec les entreprises du secteur de l'IA.
- Dans le cadre de chaque partenariat international en matière d'IA, mandater un point de contact dédié à la gestion des risques au sein du gouvernement et de l'entreprise d'IA

partenaire afin de garantir un mécanisme de responsabilité clair et un plan d'urgence en cas d'éventuelles perturbations systémiques liées à l'IA.

- Faire de la cybersécurité une priorité, en mettant l'accent sur la protection des systèmes nationaux essentiels contre les cybermenaces de plus en plus sérieuses fondées sur l'IA.

Recommandations pour les instituts ou réseaux de sécurité de l'IA

- Mesurer objectivement[22] les niveaux de sécurité et de risque liés à l'IA avant d'en généraliser la diffusion.
- Mettre en place un processus de partage des modèles de menaces pour la société et des mesures d'atténuation avec d'autres instituts ou réseaux.
- Rédiger des rapports scientifiques trimestriels sur les résultats et les bonnes pratiques afin d'informer le grand public et l'équipe chargée de la rédaction du rapport scientifique sur les connaissances acquises par les différents instituts et réseaux de sécurité.
- Encourager la recherche publique sur la sécurité de l'IA en créant un fonds mondial de recherche sur la sécurité de l'IA et un organisme de coordination chargé de définir des programmes et des standards de recherche communs.
- Soutenir et mener des recherches fondamentales publiques sur les risques systémiques en s'appuyant sur la classification européenne.[23]

Notes de fin

1 The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. (2023, 11 1). Récupéré 11 25, 2024, depuis https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

2 World leaders, top AI companies set out plan for safety testing of frontier as first global AI Safety Summit concludes. (2023, 11 2). Récupéré 11 25, 2024, depuis https://www.gov.uk/government/news/world-leaders-top-ai-companies-set-out-plan-for-safety-testing-of-frontier-as-first-global-ai-safety-summit-concludes

3 DSIT, & AI Safety Institute. (2024, 5 17). International Scientific Report on the Safety of Advanced AI. Récupéré 11 25, 2024, depuis https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai

4 DSIT. (2024, 5 21). Frontier AI Safety Commitments, AI Seoul Summit 2024. Récupéré 11 25, 2024, depuis https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024

5 U.S. Department of Commerce. (2024, 5 21). U.S. Secretary of Commerce Gina Raimondo Releases Strategic Vision on AI Safety, Announces Plan for Global Cooperation Among AI Safety Institutes. Récupéré 11 25, 2024, depuis https://www.commerce.gov/news/press-releases/2024/05/us-secretary-commerce-gina-raimondo-releases-strategic-vision-ai-safety

6 Future of Life Institute. (n.d.). EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act. Récupéré 11 25, 2024, depuis https://artificialintelligenceact.eu/

7 Future of Life Institute. (n.d.). Strengthening the European AI Act. Récupéré 11 25, 2024, depuis https://futureoflife.org/project/eu-ai-act/

8 United Nations. (2024). Governing AI for Humanity: Final Report. 9789211067873.

9 U.S. Department of Commerce. (2024, 09 18). U.S. Secretary of Commerce Raimondo and U.S. Secretary of State Blinken Announce Inaugural Convening of International Network of AI Safety Institutes in San Francisco. Récupéré 11 25, 2024, depuis https://www.commerce.gov/news/press-releases/2024/09/us-secretary-commerce-raimondo-and-us-secretary-state-blinken-announce

10 International Network of AI Safety Institutes. (2024, 11 21). Mission Statement. Récupéré 11 25, 2024, depuis https://www.nist.gov/system/files/documents/2024/11/20/Mission%20Statement%20-%20International%20Network%20of%20AISIs.pdf

11 Department for Science, Innovation and Technology (2025). International AI Safety Report 2025. [en ligne] GOV.UK. depuis https://www.gov.uk/government/publications/international-ai-safety-report-2025

12 Élysée. (n.d.). Sommet pour l'action sur l'Intelligence Artificielle. Récupéré 11 25, 2024, depuis https://www.elysee.fr/sommet-pour-l-action-sur-l-ia

13 NIST. (2024, 11 20). FACT SHEET: U.S. Department of Commerce & U.S. Department of State Launch the International Network of AI Safety Institutes at Inaugural Convening in San Francisco. Récupéré 11 25, 2024, depuis https://www.nist.gov/news-events/news/2024/11/fact-sheet-us-department-commerce-us-department-state-launch-international

14 Élysée. (2024, 5 6). Déclaration conjointe entre la République française et la République populaire de Chine sur l'intelligence artificielle et la gouvernance des enjeux globaux. Récupéré 11 25, 2024, depuis https://www.elysee.fr/emmanuel-macron/2024/05/06/declaration-conjointe-entre-la-republique-francaise-et-la-republique-populaire-de-chine-sur-lintelligence-artificielle-et-la-gouvernance-des-enjeux-globaux

15 Concordia AI. (2024, 07 19). Concordia AI holds the Frontier AI Safety and Governance Forum at the World AI Conference. Récupéré 11 25, 2024, depuis https://aisafetychina.substack.com/p/concordia-ai-holds-the-frontier-ai

16 Concordia AI. (2024, 8 2). What does the Chinese leadership mean by "instituting oversight systems to ensure the safety of AI?". Récupéré 11 25, 2024, depuis https://aisafetychina.substack.com/p/what-does-the-chinese-leadership

17 Oxford Martin School (2025). The Future of the AI Summit Series. [en ligne] Oxford Martin School. Récupéré 2 5, 2025, depuis https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-the-ai-summit-series

18 The Future Society. (2024, 11 6). Delivering Solutions: An Interim Report of Expert Insights for France's 2025 AI Action Summit. Récupéré 11 25, 2024, depuis https://thefuturesociety.org/consultation-interim-report

19 National Technical Committee 260 (TC260). (2024, 9). AI Safety Governance Framework. Récupéré 11 25, 2024.

20 Sheehan, M. (2023, 7 10). China's AI Regulations and How They Get Made. Récupéré 11 25, 2024, depuis https://carnegieendowment.org/research/2023/07/chinas-ai-regulations-and-how-they-get-made?lang=en

21 Divers auteurs. (2024, 10 3). Sécurité de l'IA Multilingue: Une lettre ouverte. Récupéré 11 25, 2024, depuis https://securitemultilingue.ai/

22 Dalrymple, D., Skalse, J., Bengio, Y., Russell, S., & Tegmark, M. (2024, 7 8). Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems. arXiv. https://doi.org/10.48550/arXiv.2405.06624.

23 European Commission. (2024, 11 14). First Draft of the General-Purpose AI Code of Practice published, written by independent experts | Shaping Europe's digital future. Récupéré 11 25, 2024, depuis https://digital-strategy.ec.europa.eu/en/library/first-draft-general-purpose-ai-code-practice-published-written-independent-experts

Recommandations de FLI pour le Sommet pour l'Action sur l'IA

Institut pour le Futur de la Vie (FLI)

Consulter ce document en ligne :

futureoflife.org/fr/document/sommet-2025/