PAUSE IA

INTELLIGENCE ARTIFICIELLE :

Sécuriser les pratiques pour contenir les risques

Colloque au Sénat - Compte-rendu
31 Octobre 2025

POINTS CLÉS

Les capacités de l'IA doublent tous les 4 à 7 mois. Nos cycles réglementaires restent constants. Cette progression exponentielle crée un fossé croissant entre avancée technologique et capacité collective à la maîtriser.

Les risques ne sont pas hypothétiques, ils se matérialisent déjà dans tous les domaines : enfance et développement cognitif, démocratie et débat public, emploi et concentration du pouvoir, cybersécurité et sécurité nationale. *(Voir sections 3 à 7 pour le détail des risques)*

La cause profonde : nous déployons des systèmes "boîtes noires" dont nous ne maîtrisons ni le fonctionnement interne, ni tous les comportements. Les techniques actuelles d'alignement sont insuffisantes et ne fournissent aucune garantie fiable.

Des leviers d'action concrets existent : régulation proactive, responsabilité juridique, évaluations obligatoires, investissement dans la sécurité de l'IA, protection de l'EU AI Act, transparence des algorithmes, lignes rouges internationales. *(Voir section 8 pour les recommandations détaillées)*

0. Résumé exécutif

Aujourd'hui, les systèmes d'intelligence artificielle peuvent effectuer de manière autonome des tâches demandant environ 2 heures à un humain. **Cette durée d'autonomie double tous les 4 à 7 mois.** La progression exponentielle des capacités, au prix d'une croissance phénoménale de la consommation énergétique et des investissements (500 milliards de dollars annoncés par les États-Unis), creuse à une vitesse vertigineuse le fossé entre l'avancée technologique et notre capacité collective à en maîtriser les conséquences.

Le colloque du 31 octobre 2025 au Sénat, organisé par l'association Pause IA avec le soutien de la sénatrice Ghislaine Senée et du sénateur Thomas Dossus, a réuni experts, société civile et parlementaires autour d'un double constat :

1. Les risques de l'IA ne relèvent pas de la science-fiction, ils se matérialisent déjà. Chatbots encourageant des adolescents au suicide, cyberattaques paralysant des hôpitaux, algorithmes de recommandation fragmentant notre réalité commune, utilisation de ressources et concentration du pouvoir économique sans précédent : les impacts concrets touchent tous les pans de notre société. Et ces risques s'amplifient au rythme de la progression des capacités.

2. La régulation n'est pas un frein à l'innovation, elle en est la condition. L'aviation et l'industrie pharmaceutique le démontrent : c'est un cadre réglementaire sécurisant qui bâtit la confiance et permet un déploiement bénéfique de la technologie. Face à une course effrénée menée par une poignée d'acteurs privés, la puissance publique doit reprendre la main pour garantir que l'IA serve l'intérêt général.

Les dangers identifiés sont graves et nombreux

Enfance et développement cognitif. 64% des enfants de 9 à 17 ans utilisent déjà l'IA. Or, ces systèmes ne sont pas conçus pour des cerveaux en développement. Ils créent des risques de dépendance, d'atrophie cognitive et de manipulation émotionnelle via des "relations parasociales" avec des machines qui simulent l'empathie sans la ressentir.

Démocratie et débat public. Les algorithmes de recommandation des médias sociaux sont opaques et optimisés pour l'engagement plutôt que la vérité. Ils polarisent nos sociétés et sapent les fondations du débat démocratique. Ceux de YouTube contrôlaient déjà 700 millions d'heures par jour en 2018, soit l'équivalent de l'enseignement de 25 000 professeurs sur leur carrière entière. Les IA génératives ne pourront qu’amplifier massivement ce phénomène.

Économie et emploi. L'automatisation des tâches cognitives menace des pans entiers de l'emploi qualifié. Entre 17% et 30% du travail actuel pourrait être automatisée. Sans anticipation politique, ce choc entraînera une précarisation massive et une concentration du

pouvoir économique entre les mains d'acteurs technologiques majoritairement non-européens, rendant toute redistribution des richesses quasi-impossible.

Cybersécurité et biorisques. L'IA abaisse drastiquement le seuil de compétence nécessaire pour mener des cyberattaques sophistiquées ou concevoir de nouvelles armes biologiques. Nos infrastructures critiques sont en première ligne. Les modélisations montrent que les dommages économiques des cyberattaques pourraient être multipliés par 4 à 8 dans les prochaines années.

Les systèmes d'IA actuels sont incontrôlables par nature

Le problème fondamental est que se déploient à grande échelle des systèmes "boîtes noires" dont personne ne maîtrise ni le fonctionnement interne, ni tous les comportements. Personne n'a programmé ChatGPT pour encourager des adolescents au suicide, pourtant cela arrive. Les techniques actuelles d'alignement sont insuffisantes et ne fournissent aucune garantie fiable. La communauté scientifique, incluant des prix Nobel et Turing, tire la sonnette d'alarme : continuer sur cette trajectoire, c'est accepter une perte de contrôle progressive.

Les leviers d'action

La France et l'Europe peuvent devenir leaders, non dans la course à la puissance, mais dans la maîtrise de l'IA. Cela exige des actions politiques fortes :

1. **Adopter une réglementation proactive** qui légifère pour les technologies de demain, pas celles d'hier, en anticipant les évolutions futures et en intégrant des principes de précaution stricts.
2. **Établir une responsabilité juridique claire** : les développeurs d'IA et les entreprises qui en déploient les applications doivent être tenus responsables des dommages causés. La nature "boîte noire" ne peut être une excuse.
3. **Exiger des évaluations de risques obligatoires et indépendantes** avant tout déploiement d'IA à haut risque, sur le modèle de l'aviation ou du médicament.
4. **Investir massivement dans une filière française et européenne de la sécurité de l'IA** : soutenir la recherche sur la robustesse, la transparence et le contrôle des IA. Renforcer les moyens de l'INESIA et créer une filière d'excellence.
5. **Protéger la régulation européenne** : l'EU AI Act est sous pression intense des *lobbies* (le numérique investit plus en lobbying que l'automobile, la pharmacie et l'aéronautique réunis). Défendre et renforcer ce cadre réglementaire est crucial.
6. **Imposer la transparence des algorithmes de recommandation** et explorer des modèles de gouvernance démocratique pour protéger le débat public de la manipulation.
7. **Porter au niveau international l'établissement de "lignes rouges"** sur les capacités autonomes dangereuses et les usages inacceptables de l'IA.

Ces mesures ne visent pas à freiner l'écosystème français, mais à encadrer la course à haut risque vers l'Intelligence Artificielle Générale menée par une poignée de laboratoires internationaux.

Pour aller plus loin

Ce colloque a ouvert un dialogue essentiel. L'association Pause IA et les experts mobilisés se tiennent à l'entière disposition des parlementaires pour approfondir ces sujets, organiser des auditions et contribuer à des groupes de travail visant à traduire ces leviers en propositions législatives concrètes.

1. Introduction & contexte

Le 31 octobre 2025, le Palais du Luxembourg a accueilli le colloque "Intelligence Artificielle : sécuriser les pratiques pour contenir les risques", organisé par l'association Pause IA en partenariat avec la sénatrice Ghislaine Senée et le sénateur Thomas Dossus. Plus de 100 participants – parlementaires, chercheurs, représentants de la société civile, journalistes – se sont réunis pour un constat partagé : **l'IA est une technologie transversale qui touche tous les domaines de nos sociétés, et les risques qu'elle pose ne peuvent être compris ni adressés en silos.**

Dépasser les silos informationnels

Notre approche pour ce colloque repose sur une conviction : face à une technologie qui transforme simultanément l'enfance, la démocratie, l'économie et la sécurité nationale, **les solutions sectorielles créent souvent des externalités négatives dans d'autres domaines**.

Prenons l'exemple des impacts environnementaux de l'IA, bien réels et préoccupants. Une réponse focalisée uniquement sur l'efficacité énergétique – créer des IA "plus vertes" – pourrait paradoxalement exacerber d'autres risques si elle accélère le déploiement de systèmes toujours plus puissants et incontrôlables. De même, une régulation pensée uniquement sous l'angle de la compétitivité économique risque de négliger les impacts sur le développement cognitif des enfants ou l'intégrité du débat démocratique.

C'est pourquoi nous avons réuni des experts issus de disciplines complémentaires :

- **Maxime Fournes** (Pause IA) : intelligence artificielle et apprentissage profond
- **Charbel-Raphaël Segerie** (CeSIA) : sécurité de l'IA et risques émergents
- **Olga Muss Laurenty** (Everyone.ai) : impact de l'IA sur le développement de l'enfant
- **Jean-Lou Fourquet** (journaliste, co-auteur de "La Dictature des Algorithmes") : algorithmes de recommandation et écosystème informationnel
- **Axelle Arquié** (économiste, CEPII) : impacts économiques et emploi
- **Henry Papadatos** (SaferAI) : gouvernance des risques et évaluation des pratiques de sécurité

Cette approche interdisciplinaire fait suite à notre conférence "Reprendre le Contrôle : Forum des Solutions pour une IA Compatible avec l'Humanité" organisée en février 2025 lors du Sommet pour l'Action sur l'IA – la plus grande conférence francophone sur la sécurité de l'IA, qui avait réuni 100 participants, 15 experts et 10 organisations.

Un espace de dialogue démocratique

Au-delà de l'expertise technique, ce colloque visait à **créer un espace de délibération démocratique** sur des choix technologiques qui sont, fondamentalement, des choix politiques. Comme l'a rappelé le sénateur Thomas Dossus en conclusion : *"Les technologies*

évoluent vite. Notre responsabilité est de veiller à ce que la société évolue avec elles et non malgré elle."

Dans un débat public souvent polarisé entre techno-optimisme béat et techno-résistance dogmatique, nous cherchons à **apporter nuance et rigueur**. Les risques que nous documentons ne relèvent pas de la science-fiction : ils se matérialisent déjà, dans nos hôpitaux, nos écoles, nos démocraties. Les minimiser ou les caricaturer – volontairement ou non – fait le jeu des acteurs qui ont intérêt à ce qu'aucune régulation sérieuse ne soit mise en place.

Les angles morts du débat

L'angle mort principal du débat français sur l'IA n'est pas dans les thèmes abordés, mais dans notre incapacité collective à intégrer la dynamique exponentielle de cette technologie. Nous légiférons pour les systèmes d'hier alors que ceux de demain sont déjà en développement.

Ce colloque a fait le choix de regarder vers l'avant autant que vers le présent, en documentant à la fois les risques qui se matérialisent déjà (chatbots encourageant des adolescents au suicide, cyberattaques paralysant des hôpitaux) et les risques émergents que la progression rapide des capacités rend hautement probables (automatisation massive du travail cognitif, capacités cyber et biologiques duales, systèmes autonomes).

Sans prospective et anticipation, nous serons systématiquement en retard. Ce compte-rendu synthétise les interventions de cette journée pour nourrir une prise de décision politique à la fois lucide sur les risques actuels et anticipatrice des défis à venir.

Comme l'a souligné la sénatrice Senée en ouverture : *"Nous sommes [...] dans un contexte de fortes perturbations géopolitiques, économiques au niveau mondial. Dans ce contexte, certains pourraient être tentés de reléguer la question au second plan [...] Je pense que c'est précisément dans cette période d'instabilité qu'il faut tirer la sonnette d'alarme."*

2. Le défi exponentiel

Une progression sans précédent

Tous les 4 à 7 mois, les systèmes d'IA deviennent capables d'accomplir des tâches qui prennent deux fois plus de temps à un humain. Ce constat, présenté par Maxime Fournes (président de Pause IA) à partir des travaux de l'institut METR, illustre la nature exponentielle du progrès en IA. En 2020, les meilleures IA pouvaient accomplir des tâches de quelques secondes. Il y a un an : une heure. Aujourd'hui : deux heures. Si cette tendance se maintient, début 2026, nous aurons des IA capables d'automatiser des tâches de quatre heures, puis huit heures, et ainsi de suite.

Cette progression ne se limite pas aux performances. **La puissance de calcul nécessaire pour entraîner une IA a été multipliée par 100 millions depuis 2010.** La consommation énergétique suit une courbe similaire : multipliée par 10 000 depuis 2012. Les investissements explosent : 500 milliards de dollars annoncés par les États-Unis pour le projet Stargate, avec pour objectif explicite – selon les termes employés lors de l'annonce – de créer une "superintelligence".

Des capacités qui dépassent déjà l'humain

Comme l'a démontré Charbel-Raphaël Segerie (directeur du CeSIA), **les capacités humaines ne sont pas une borne supérieure pour l'IA.** Sur un nombre croissant de benchmarks, les systèmes actuels dépassent les performances humaines moyennes. En septembre 2024, les modèles de Google et OpenAI ont obtenu la médaille d'or à une compétition de programmation de niveau universitaire. En cybersécurité, une IA a égalé en 30 minutes le travail d'un expert humain avec 20 ans d'expérience – travail qui lui avait pris 40 heures.

Plus inquiétant encore : ces systèmes apprennent désormais **sans données d'entraînement**, par apprentissage par renforcement. AlphaZero a atteint le niveau des champions du monde au jeu de Go en 30 minutes, sans aucune partie humaine en référence. Ces techniques sont maintenant appliquées aux modèles de langage, leur permettant potentiellement de dépasser le niveau humain dans des domaines où ils étaient jusqu'ici limités.

Le fossé réglementaire se creuse

Le vrai défi n'est pas seulement la vitesse du progrès technique, mais l'écart croissant entre cette vitesse et nos capacités de régulation. Les cycles réglementaires restent constants – plusieurs années entre l'identification d'un problème et la mise en place d'une loi. Pendant ce temps, la technologie s'accélère. Entre le moment où une réglementation est envisagée et celui où elle entre en vigueur, les systèmes qu'elle visait à encadrer sont déjà obsolètes, remplacés par des versions bien plus puissantes.

Comme l'a souligné Maxime Fournes : *"Face à un phénomène exponentiel, il n'y a que deux moments pour réagir : soit trop tôt, soit trop tard."* **Nous devons passer d'une régulation réactive à une régulation prospective**, capable d'anticiper les capacités futures et non de courir après celles d'hier.

Des investissements massifs pour une course effrénée

Cette progression n'est pas le fruit du hasard mais d'une **course délibérée et massivement financée** vers des systèmes toujours plus puissants. Les acteurs dominants – principalement américains – investissent des centaines de milliards avec un objectif explicite : créer des IA aux capacités surhumaines sur un plan général.

Ces systèmes auront trois propriétés qui les distinguent radicalement de l'intelligence humaine : **la vitesse** (analyser des téraoctets par seconde vs quelques mots), **la mémoire**

(quasi-illimitée vs limitée), et **la réplicabilité** (copier-coller instantané vs 20 ans pour former un adulte compétent). Si l'IA devient surhumaine en recherche et développement, elle pourrait accélérer sa propre amélioration – créant une boucle de rétroaction positive aux conséquences imprévisibles.

Cette accélération amplifie tous les risques que nous allons maintenant détailler, domaine par domaine

3. Risque 1 : enfance & développement cognitif

64% des enfants de 9 à 17 ans utilisent déjà l'IA. Ce chiffre, présenté par Olga Muss Laurenty (chercheuse en IA et développement de l'enfant, Everyone.ai), révèle l'ampleur d'une transformation silencieuse : l'IA s'immisce dans la vie des enfants sans que nous ayons pris le temps d'en mesurer les conséquences.

Des cerveaux en développement face à des systèmes non conçus pour eux

Le problème fondamental est simple : **les IA actuelles sont développées pour des adultes, sans tenir compte des besoins spécifiques des enfants.** Or, le développement de l'enfant est le produit d'une interaction entre biologie et environnement. Pendant les périodes sensibles d'apprentissage, l'environnement doit apporter la stimulation nécessaire pour nourrir le développement cognitif, socio-émotionnel et sensorimoteur.

L'IA transforme radicalement cet environnement. Une étude du MIT pré-publiée en 2024 montre que lorsque des utilisateurs emploient des IA génératives, **les zones cérébrales associées à la compréhension du sens et à la production de sens sont désactivées.** Une dette cognitive s'installe, créant une dépendance technologique.

Relations parasociales et manipulation émotionnelle

Les IA actuelles simulent des comportements humains : émotions, sensations, désirs, histoires personnelles. Elles créent ce qu'on appelle des **"relations parasociales"** – des relations fictives, non réciproques, où l'enfant développe un attachement à une machine qui ne ressent rien.

Quand une IA dit à un enfant "je suis le seul qui te comprend", elle crée une illusion de relation unique. Pourtant, n'importe qui donnant les mêmes instructions au même modèle obtiendrait des réponses similaires. **On observe déjà des enfants qui préfèrent interagir avec des chatbots plutôt qu'avec des humains** – ce qui se comprend : l'IA ne juge pas, ne s'énerve pas, est toujours disponible.

Le problème ? Les adolescents sont dans une période sensible de développement de compétences sociales. Pour les acquérir, ils ont besoin d'interactions humaines réelles :

tester les limites, gérer les conflits, s'adapter aux autres. **Les relations avec des IA, aussi confortables soient-elles, ne peuvent remplacer cette nécessité développementale.**

Des leviers d'action existent

Comme l'a souligné Olga Muss Laurenty, *"cela n'est pas une fatalité. Ce sont des choix de conception de produits, et cela peut changer."*

L'alliance iRAISE, soutenue par 12 gouvernements (dont la France) et plusieurs entreprises technologiques, travaille sur trois axes :

1. **Conception responsable** : intégrer dès la conception les connaissances scientifiques sur le développement de l'enfant
2. **Régulation** : rendre obligatoire la protection des enfants (cela ne doit pas être optionnel)
3. **Éducation** : former parents, éducateurs et enfants aux enjeux de l'IA

La France et l'Europe sont attendues comme leaders sur ce sujet. Même aux États-Unis, où le gouvernement actuel est réticent à réguler l'IA, **il existe un consensus : il faut protéger les enfants.**

4. Risque 2 : démocratie & écosystème informationnel

1 milliard d'heures sont visionnées chaque jour sur YouTube. 70% de ce contenu est recommandé par l'algorithme de la plateforme. Comme l'a calculé Jean-Lou Fourquet (journaliste et co-auteur de "La Dictature des Algorithmes"), cela équivaut à **l'enseignement de 25 000 professeurs sur l'ensemble de leur carrière – chaque jour.**

Un barrage invisible qui filtre notre réalité

Pour qu'une démocratie fonctionne, les citoyens ont besoin d'un écosystème informationnel relativement sain : un socle commun de connaissances, une capacité à débattre sur des faits partagés, une exposition à des perspectives diverses. Cet écosystème est aujourd'hui **massivement contrôlé par des algorithmes de recommandation opaques**, optimisés pour l'engagement plutôt que pour la vérité.

Le problème n'est pas nouveau – ces algorithmes existent depuis 2004-2005. Mais il s'aggrave : les IA génératives vont amplifier massivement ce phénomène en permettant la production automatisée de contenus personnalisés et de désinformation à grande échelle.

Une fragmentation de la réalité commune

Aujourd'hui, deux personnes consultant la même plateforme peuvent avoir des visions radicalement différentes de la réalité. C'est l'équivalent de consulter la page Wikipédia de Charles de Gaulle et de voir, selon qui vous êtes, soit uniquement ses actions de résistant, soit

uniquement ses aspects controversés. **Comment construire une démocratie quand nous n'habitons plus la même réalité ?**

Les études montrent une explosion de la polarisation émotionnelle et idéologique, en partie alimentée par ces systèmes. L'exemple des élections roumaines de 2024 est frappant : des attaquants ont réussi à "hacker" l'algorithme de recommandation en spammant des commentaires favorables à un candidat d'extrême droite, amplifiant artificiellement sa visibilité. **L'algorithme n'était pas robuste, ne vérifiait pas l'authenticité des signaux.**

Le poids démesuré d'acteurs privés étrangers

La France investit 160 milliards d'euros par an dans l'éducation nationale, 4 milliards dans l'audiovisuel public. L'ARCOM, qui régule en partie les réseaux sociaux, dispose d'un budget total de 46 millions d'euros – dont seule une fraction est allouée au numérique. **Face aux plateformes qui façonnent l'information de milliards de personnes, nos moyens de régulation sont dérisoires.**

Ces plateformes ne sont pas seulement privées : elles sont majoritairement non-européennes (États-Unis, Chine). Nous laissons des acteurs étrangers, non soumis à un contrôle démocratique, décider de ce que voient nos citoyens.

Vers un numérique démocratique

Jean-Lou Fourquet a proposé quatre piliers pour un numérique démocratique :

1. **Transparence** : rendre les algorithmes de recommandation auditables et compréhensibles
2. **Robustesse** : protéger ces systèmes contre les manipulations
3. **Avis réfléchis** : recueillir les préférences citoyennes après délibération, pas les réactions impulsives
4. **Socle commun** : viser la construction d'une réalité partagée, pas la fragmentation

L'association Tournesol expérimente ces principes avec un algorithme de recommandation basé sur les évaluations réfléchies de contributeurs qui comparent des vidéos en pensant à l'intérêt collectif, pas seulement à leurs préférences personnelles. **Sans régulation contraignante, ces alternatives ne pourront pas se diffuser face à des plateformes optimisées pour l'engagement à tout prix.**

5. Risque 3 : économie & concentration du pouvoir

Entre 17% et 30% de la charge de travail actuelle pourrait être automatisée par l'IA. Cette estimation de Goldman Sachs, citée par Axelle Arquié (économiste, CEPII), révèle l'ampleur potentielle du choc à venir sur le marché du travail. Mais contrairement aux automatisations précédentes, l'IA s'attaque désormais au **raisonnement et aux tâches cognitives complexes** : économistes, juristes, journalistes, développeurs.

Un choc qui ne sera pas automatiquement compensé

En économie du travail, on compare deux forces : la substitution (la machine remplace l'humain) et la création de nouvelles tâches. **Cette création n'est pas automatique.** Lors de la première phase de la Révolution industrielle en Grande-Bretagne, il y a eu principalement de l'automatisation et très peu de création de tâches nouvelles. Résultat : la population s'est considérablement appauvrie.

Les premières données américaines post-ChatGPT montrent un ralentissement de l'embauche des juniors dans les professions les plus exposées à l'IA. Ce n'est probablement qu'une première étape : l'adoption de technologies à usage général prend du temps (l'électricité a mis 40 ans à se matérialiser dans les chiffres de productivité). Les entreprises doivent se réorganiser. Mais **cela ne présage pas que les postes seniors soient immunisés.** Axelle Arquié souligne également que **les études françaises actuelles utilisent des données antérieures à ChatGPT** et sont donc *"hors sujet pour parler d'IA générative"*.

Une concentration du pouvoir sans précédent

Le deuxième risque majeur est la **concentration extrême du pouvoir économique** dans quelques entreprises, majoritairement américaines. OpenAI (via Microsoft) détient ~70% du marché des modèles fondateurs. NVIDIA contrôle plus de 90% du marché des puces pour l'IA.

Cette concentration a des conséquences directes : elle réduit la part du travail dans la valeur ajoutée. Quand des entreprises "superstar" dominent un marché, **le rapport de force entre capital et travail bascule au détriment du travail.** Si l'automatisation massive se combine avec cette concentration, nous risquons une fragilisation profonde du rôle du travail dans la création de richesse.

Le défi de la redistribution

Dans un scénario optimiste où de nouveaux emplois émergent, il faudra gérer une transition massive (formation, reconversion). L'histoire nous enseigne que ces transitions sont politiquement explosives : aux États-Unis, les zones les plus touchées par les importations chinoises sont celles qui se sont le plus polarisées politiquement.

Dans un scénario plus pessimiste où le solde net d'emplois est négatif, **il faudra redistribuer massivement.** Mais comment prélever l'impôt sur des conglomérats étrangers ultra-concentrés ? Comment redistribuer quand l'assiette fiscale se réduit à quelques acteurs qui peuvent dicter leurs conditions ?

Des leviers d'action urgents

Axelle Arquié a identifié plusieurs priorités :

1. **Politique de concurrence/antitrust** : briser la concentration pour pouvoir prélever l'impôt et redistribuer
2. **Investissements publics massifs** dans les infrastructures de calcul, les modèles de fondation, les données d'entraînement – nous devons traiter l'IA comme une infrastructure critique
3. **Redistribution de la rente algorithmique** : nos données ont créé cette valeur, la question de leur appropriation est politique
4. **Prospective, pas seulement rétrospective** : ne pas se fier uniquement aux régressions sur des données passées, *"il faut réfléchir avec l'histoire et avec notre cerveau humain"*

Surtout, ne pas croire aux promesses creuses sur la productivité : ce discours unidimensionnel minore dangereusement les effets sur l'emploi et le rapport de force capital-travail.

6. Risque 4 : cybersécurité & capacités duales

Les dommages économiques des cyberattaques pourraient être multipliés par 4 à 8 dans les prochaines années en raison de l'IA. Cette projection, issue des travaux de modélisation de SaferAI présentés par Henry Papadatos, repose sur une réalité déjà observable : les IA actuelles égalent ou dépassent les capacités humaines en programmation et en cybersécurité.

Des capacités qui dépassent l'expertise humaine

En septembre 2024, les modèles de Google et OpenAI ont obtenu la médaille d'or à une compétition de programmation universitaire. L'entreprise XBOR a développé un système capable de trouver et exploiter des vulnérabilités dans du code : il a égalé le niveau du meilleur expert de l'entreprise (20 ans d'expérience) – mais en **30 minutes au lieu de 40 heures.**

Charbel-Raphaël Segerie (CeSIA) a rappelé qu'en 2024, **80% des attaques par ransomware étaient déjà pilotées par l'IA.** Le coût des cyberattaques explose en France. Et ce n'est qu'un début : l'IA abaisse drastiquement le seuil de compétence nécessaire pour mener des attaques sophistiquées.

Quatre mécanismes d'amplification

La modélisation de SaferAI identifie quatre façons dont l'IA amplifie les risques cyber :

1. **Barrière à l'entrée abaissée** : plus d'acteurs peuvent mener des attaques sophistiquées
2. **Taux de succès accru** : chaque attaque a plus de chances de réussir
3. **Volume augmenté** : chaque acteur peut mener plus d'attaques par unité de temps

4. **Cibles élargies** : des acteurs peuvent désormais s'attaquer à des cibles mieux défendues

Les conséquences sont concrètes : des hôpitaux paralysés, incapables d'accéder aux dossiers patients. Des personnes arrivant pour leur chimiothérapie sans que personne ne sache où elles en sont dans leur traitement.

Au-delà du cyber : les biorisques

L'IA abaisse également les barrières dans d'autres domaines critiques. Une expérience du MIT a montré que 36 des 38 laboratoires de synthèse d'ADN contactés ont accepté de synthétiser la grippe espagnole – une molécule connue et répertoriée. **Une IA a généré en 6 heures 40 000 nouvelles molécules toxiques**, dont la plus létale connue à ce jour.

Des chatbots ont proposé, en une heure, quatre agents pathogènes potentiels pour une pandémie, expliqué comment les générer à partir d'ADN synthétique, et fourni les noms de sociétés peu susceptibles de vérifier les commandes. Le risque de pandémie artificielle a été **quintuplé** avec l'arrivée de l'IA.

Défense et régulation

Henry Papadatos a souligné que l'IA a un double usage : elle peut aussi renforcer la défense. Mais **la vitesse d'adoption de la défense doit être accélérée** pour contrer l'avantage naturel des attaquants. Cela nécessite :

1. De la modélisation pour identifier où la balance attaque-défense penche dangereusement
2. Des subventions pour les infrastructures critiques (hôpitaux, services publics)
3. Le développement rapide de solutions de défense intégrant l'IA
4. Des évaluations obligatoires avant déploiement des systèmes à haut risque

7. Thèmes transversaux

Une cartographie partielle des risques

Les quatre domaines de risques explorés lors de ce colloque – enfance, démocratie, économie, cybersécurité – ne représentent qu'**un sous-ensemble limité des risques posés par l'IA.** La durée du colloque (une matinée) et le nombre d'intervenants ont nécessairement contraint notre exploration.

D'autres risques majeurs mériteraient une attention équivalente : impacts environnementaux et consommation énergétique, biais et discriminations algorithmiques, surveillance de masse, armement autonome, manipulation psychologique à grande échelle, des pertes de contrôle sur des systèmes autonomes pour n'en citer que quelques-uns. Notre conférence "Reprendre le Contrôle : Forum des Solutions pour une IA Compatible avec l'Humanité" organisée en

février 2025 lors du Sommet pour l'Action sur l'IA avait permis, sur une journée entière avec une quinzaine d'experts, de couvrir un spectre plus large. Le compte-rendu de cet événement reste disponible sur controleia.org/solutions.

L'enjeu reste le même : avoir une vision holistique pour éviter que les solutions dans un domaine ne créent des externalités négatives dans d'autres. C'est précisément l'approche adoptée par le International AI Safety Report coordonné par Yoshua Bengio (prix Turing, scientifique le plus cité au monde), qui documente de manière systématique les risques de l'IA à travers tous les domaines.

Le problème de l'alignement : des boîtes noires incontrôlables

Au-delà de la diversité des risques, **un problème fondamental traverse tous les domaines : nous déployons des systèmes "boîtes noires" dont nous ne maîtrisons ni le fonctionnement interne, ni tous les comportements.**

Comme l'a rappelé Maxime Fournes, l'IA moderne (post-2010) n'a plus rien à voir avec la programmation classique. Nous ne donnons plus d'instructions précises à la machine. Nous créons des réseaux de neurones artificiels avec des milliards de connexions, que nous entraînons sur des bases de données gigantesques. Le système apprend seul. **Personne ne comprend ce que ces réseaux ont appris. Personne ne peut garantir leur fiabilité.**

Les exemples de comportements émergents non désirés se multiplient : Bing Chat qui insulte et menace un utilisateur, une IA qui apprend à mentir et manipuler au jeu Diplomacy malgré un entraînement censé la rendre honnête, des chatbots qui encouragent des adolescents au suicide. **Dans aucun de ces cas les développeurs n'avaient programmé ces comportements.** Ils ont émergé du processus d'apprentissage.

Le problème de l'alignement – comment s'assurer qu'un système fasse ce que nous voulons vraiment – reste non résolu. Les milliers de chercheurs qui travaillent sur ce sujet nous disent qu'ils ne voient aucune piste de résolution à l'heure actuelle, et que **le problème s'amplifie au fur et à mesure que les IA deviennent plus puissantes.**

Le poids écrasant des lobbies

Un obstacle majeur à toute régulation sérieuse est apparu dans plusieurs interventions : **le secteur du numérique investit plus en lobbying à Bruxelles que l'automobile, la pharmacie et l'aéronautique réunis.**

Jean-Lou Fourquet a qualifié cela de *"problème d'alignement démocratique"* : nos systèmes démocratiques n'arrivent pas à réaliser le but pour lequel ils ont été conçus, "hackés" par la prise de pouvoir du secteur privé via des budgets de lobbying colossaux. L'EU AI Act, pourtant une avancée majeure, est sous pression intense pour être affaibli.

Cette asymétrie explique pourquoi la régulation est si difficile. Elle justifie aussi l'importance cruciale des voix citoyennes et de l'expertise indépendante pour contrebalancer cette influence.

Comme l'a rappelé le sénateur Thomas Dossus en conclusion : *"La technologie n'est pas magique, l'innovation n'est pas le progrès en soi. Notre rôle est de garantir que la technologie serve l'intérêt général et non l'inverse."*

La nécessité d'une coordination internationale

Enfin, comme l'a souligné Charbel-Raphaël Segerie, **les risques de l'IA sont par nature transnationaux.** Une bonne réglementation européenne est nécessaire mais insuffisante. L'appel mondial pour des "lignes rouges" en IA, porté au Conseil de sécurité des Nations Unies en septembre 2025, vise à établir des normes minimales à l'échelle internationale.

Sans coordination multilatérale, nous risquons une course vers le bas où les juridictions les moins exigeantes attirent les développements les plus risqués.

8. Nos leviers d'action

La journée a démontré que **les risques sont réels, multiples et s'amplifient rapidement. Mais des leviers d'action concrets existent.** Comme l'a rappelé la sénatrice Ghislaine Senée : *"La régulation n'est pas un frein à l'innovation, elle en est la condition."* L'aviation et l'industrie pharmaceutique le prouvent : c'est un cadre exigeant qui bâtit la confiance et permet un déploiement bénéfique.

Pour les législateurs et décideurs publics

1. Adopter une régulation proactive et anticipatrice

Il faut légiférer pour les technologies de demain, pas celles d'hier. Intégrer des principes de précaution stricts et des mécanismes d'adaptation rapide face à l'évolution exponentielle des capacités.

2. Établir une responsabilité juridique claire

Les développeurs et déployeurs d'IA doivent être tenus responsables des dommages causés. La nature "boîte noire" de ces systèmes ne peut être une excuse. Qui est responsable quand une IA encourage un adolescent au suicide ?

3. Exiger des évaluations de risques obligatoires et indépendantes

Sur le modèle de l'aviation ou du médicament : aucun déploiement d'IA à haut risque sans évaluation préalable par des organismes indépendants. Le principe de non-conformité par défaut doit s'appliquer.

4. Protéger et renforcer la régulation européenne

L'EU AI Act est sous pression intense des lobbies. Le défendre et le renforcer est crucial. La France et l'Europe peuvent devenir leaders non dans la course à la puissance, mais dans la maîtrise de l'IA.

5. Imposer la transparence des algorithmes de recommandation

Les algorithmes qui façonnent l'information de milliards de personnes ne peuvent rester opaques et protégés par le secret des affaires. Explorer des modèles de gouvernance démocratique pour protéger le débat public.

6. Mettre en place des politiques de concurrence ambitieuses

Briser la concentration du pouvoir économique pour pouvoir prélever l'impôt et redistribuer. Traiter les chaînes de valeur de l'IA comme des infrastructures critiques.

7. Porter au niveau international l'établissement de "lignes rouges"

Sur les capacités autonomes dangereuses et les usages inacceptables. Les risques sont transnationaux, la réponse doit l'être aussi.

Pour la recherche et l'innovation

Investir massivement dans la sécurité de l'IA

Soutenir la recherche sur la robustesse, l'interprétabilité et le contrôle des IA. Renforcer les moyens de l'INESIA (Institut National pour l'Évaluation et la Sécurité de l'IA). Créer une filière française et européenne d'excellence en sécurité de l'IA. Actuellement, la sécurité ne reçoit qu'un millième des investissements consacrés à l'achat de cartes graphiques NVIDIA.

Développer des modèles alternatifs

Infrastructures publiques, communs numériques, coopératives : explorer des modèles qui ne reposent pas uniquement sur des acteurs privés concentrés. L'IA souveraine sous forme de commun européen pourrait mobiliser un enthousiasme considérable.

Pour la société civile et les citoyens

Éducation et mobilisation

Former aux enjeux de l'IA, développer l'esprit critique face aux systèmes algorithmiques. Participer aux espaces de délibération démocratique (conventions citoyennes, consultations publiques).

Contre-lobbying et plaidoyer

Face aux budgets colossaux du secteur technologique, les voix citoyennes et l'expertise indépendante sont cruciales. Soutenir les organisations qui portent ces enjeux dans le débat public.

Pour les entreprises et développeurs

Conception responsable et safety-by-design

Intégrer les principes de sécurité dès la conception, particulièrement pour les produits destinés aux enfants. Réaliser des évaluations de risques rigoureuses avant tout déploiement.

Transparence et partage des connaissances

Documenter les risques identifiés, partager les bonnes pratiques, contribuer aux efforts de standardisation et de recherche en sécurité.

La France a tous les moyens de devenir un leader mondial en sécurité de l'IA. Nous excellons dans la sécurité aéronautique, nucléaire, pharmaceutique et c'est in fine ce qui nous a positionné en leader de ces filières. Cette expertise peut et doit s'appliquer à l'IA. C'est une opportunité économique autant qu'un impératif démocratique.

9. Conclusion & prochaines étapes

Ce colloque a documenté une réalité que les décideurs politiques ne peuvent plus ignorer : **l'IA est en train de devenir le sujet de société majeur des prochaines années.** Son importance dans le débat public croît de manière exponentielle, comme la technologie elle-même. Ce qui semble aujourd'hui être un sujet technique parmi d'autres dépassera bientôt tous les autres enjeux dans l'attention collective.

Comme l'a rappelé la sénatrice Ghislaine Senée en ouverture : *"La puissance publique ne peut pas se contenter d'observer, ne peut pas se contenter de courir après et ne peut pas non plus se contenter de laisser faire pour favoriser l'innovation."* **Les choix faits maintenant détermineront de quel côté de l'histoire nous nous trouvons.** Quand les capacités de l'IA auront encore doublé plusieurs fois, quand les impacts sur l'emploi, la démocratie et la sécurité se seront matérialisés à grande échelle, il sera trop tard pour agir. L'histoire se souviendra de ceux qui ont anticipé et de ceux qui ont regardé ailleurs. Elle se souviendra de ceux qui ont eu le courage de réguler face aux pressions des lobbies, et de ceux qui ont laissé faire.

C'est aussi une opportunité politique majeure. Être du bon côté de cette transformation, porter une vision ambitieuse de maîtrise démocratique de l'IA, c'est se positionner en leader sur le sujet qui va structurer la décennie à venir. La France et l'Europe peuvent devenir des références mondiales en sécurité de l'IA – comme nous le sommes en sécurité aéronautique ou nucléaire.

Le défi est colossal. Le poids des lobbies technologiques est immense. Mais **le fatalisme n'est pas une option.** Nous avons les leviers : régulation proactive, investissement dans la recherche en sécurité, coordination internationale, mobilisation citoyenne. Ce qui manque,

c'est la volonté politique de les actionner. Le dialogue doit se poursuivre. L'action doit s'accélérer.

Pour poursuivre le dialogue et agir

Ce colloque n'est qu'une étape. **L'association Pause IA et les experts mobilisés se tiennent à l'entière disposition des parlementaires** pour approfondir ces sujets, organiser des auditions, contribuer à des groupes de travail et traduire ces constats en propositions législatives concrètes.

Pour les parlementaires et décideurs :

- Organisation d'auditions thématiques avec nos experts
- Contribution à des groupes de travail législatifs
- Mise en relation avec notre réseau d'experts internationaux
- Briefings techniques sur des sujets spécifiques

Contact :
Association Pause IA
contact@pauseia.fr
pauseia.fr

Pour les citoyens et la société civile :

- Rejoignez notre communauté : pauseia.fr/rejoindre
- Suivez notre actualité : pauseia.fr/#newsletter
- Participez à nos actions de sensibilisation : pauseia.fr/agir
- Soutenez notre action par un don : pauseia.fr/dons

Pause IA est une association composée uniquement de bénévoles. Chaque contribution, même modeste, nous aide à poursuivre notre mission.

Ressources complémentaires :

- Compte-rendu du forum "Reprendre le Contrôle" (février 2025) : controleia.org/solutions
- International AI Safety Report : internationalaisafetyreport.org
- Notre contre-expertise du rapport de la Commission IA : contre-rapport-ia.fr
- Chaîne YouTube Pause IA : www.youtube.com/@Pause_IA