



AGI PREPAREDNESS INSTITUTE

Bâtir la quatrième Genève internationale

Présentation de l'AGI Preparedness Institute



Septembre 2026

1. Pourquoi la préparation à l'IA, pourquoi Genève, pourquoi maintenant?¹

L'Institut pour la préparation à l'intelligence artificielle générale est une nouvelle organisation à but non lucratif basée à Genève qui s'emploie à préparer les institutions nationales et internationales aux conséquences de l'IA avancée. Il travaille sous son nom anglais, AGI Preparedness Institute (AGIPI), dans la Genève internationale et avec ses partenaires dans le monde entier. Ce bref document d'introduction expose la raison d'être de l'Institut, ce qu'il prévoit de faire au cours de sa première année et la manière dont vous pouvez vous impliquer.

a) Pourquoi la préparation à l'IA ?

La préparation à l'IA consiste à renforcer la capacité des institutions à gérer les conséquences sociétales, économiques et géopolitiques croissantes de l'IA avancée. Cela comprend la réduction du risque de perte de contrôle et d'usage malveillant catastrophique de l'IA, les conditions d'un épanouissement humain dans une économie post-travail et la préservation de la liberté humaine à l'ère de la superpersuasion et de la supersurveillance.

Aujourd'hui, la préparation à l'IA prend sa forme la plus opérationnelle dans les politiques de mise à l'échelle responsable (Responsible Scaling Policies) des entreprises. Il s'agit d'engagements volontaires pris par des entreprises d'IA selon une logique conditionnelle "si... alors...": plus les capacités de l'IA augmentent, plus les exigences en matière d'atténuation des risques, comme le renforcement de la cybersécurité de l'entreprise, sont élevées. C'est une bonne chose. Toutefois, la responsabilité sociale des entreprises ne suffit pas :

- Les engagements de mise à l'échelle responsable des entreprises sont volontaires, et certains de leurs aspects seraient plus crédibles du point de vue de la théorie des jeux s'ils pouvaient être vérifiés et imposés.
- De nombreux préjudices liés à l'IA ne découlent pas d'un seul modèle d'IA, mais de l'usage cumulé de nombreux systèmes à l'échelle de la société.
- Pour les menaces qui ne requièrent qu'un certain niveau absolu de capacités d'IA (par opposition à une supériorité relative sur une autre IA), de puissantes forces favorisent une diffusion ouverte, quoi que fasse un acteur isolé. Cela exige de la résilience et des plans d'urgence.
- Certains risques se gèrent plus efficacement en amont ou en aval du modèle lui-même. Les entreprises qui développent des modèles d'IA ne peuvent pas réglementer les plateformes de médias sociaux, les fournisseurs d'accès à Internet ou les goulets d'étranglement propres à certains domaines, tels que les fournisseurs de synthèse d'ADN, qui transforment des fichiers numériques en matériel biologique.

¹ Traduction automatique : ce document est une traduction automatique. La version originale anglaise fait foi : agipreparedness.org/AGIPI_Introduction.pdf

Par conséquent, il faut qu'un éventail d'acteurs, des gouvernements nationaux aux institutions internationales en passant par les partenariats public-privé, intensifient de toute urgence leurs efforts pour préparer la société à l'IA avancée. Si l'Institut utilise explicitement le terme d'intelligence artificielle générale (IAG) dans son nom, c'est parce que l'IAG est un terme tourné vers l'avenir. On ne peut ni comprendre ni gouverner l'IA de manière adéquate en se contentant d'observer l'état de l'art actuel. La puissance de calcul totale du parc de puces d'IA et l'efficacité de ce calcul augmentent de façon exponentielle. L'intelligence devient environ 40 fois moins chère chaque année ! Les modèles d'IA sont passés d'un score de 5% en mathématiques de niveau secondaire en 2021 à la résolution de problèmes de recherche ouverts en mathématiques, dont le premier problème du prix du Millénaire en septembre 2026. De 17.5% sur Cybench en 2024 à la découverte d'un nombre sans précédent de vulnérabilités dans des logiciels réels en 2026. Tout ce qui se veut à l'épreuve de l'avenir doit anticiper la possibilité que l'IA continue de progresser rapidement.

Il est tout aussi important de préciser ce que l'Institut n'entend pas par préparation à l'IAG. Il ne s'agit pas de se préparer à un unique "jour de l'IAG", un événement ponctuel après lequel l'humanité aurait soit résolu tous ses problèmes, soit disparu. Selon mon modèle mental, la transition vers une "ère de l'IAG", dans laquelle les machines prendront les décisions les plus lourdes de conséquences, s'apparente davantage à une "révolution industrielle comprimée", une période de changement soutenu et accéléré plutôt qu'un seuil unique. L'objectif de la préparation est de façonner cette transition de sorte que l'humanité s'épanouisse et conserve durablement une réelle capacité d'agir.

b) Pourquoi Genève ?

Les écosystèmes d'institutions et de personnes peuvent se révéler étonnamment tenaces. Ils créent des filières de talents spécialisés et un savoir-faire tacite sur le fonctionnement réel des institutions, qui ne se diffuse qu'à l'échelle locale. Nombreux sont ceux qui ont tenté de bâtir la prochaine Silicon Valley, et pourtant ces efforts ont largement échoué. Malgré tous ses défis et toutes ses lacunes, San Francisco reste de loin la capitale mondiale de l'IA.

De la même manière, **Genève est un écosystème de gouvernance internationale tenace depuis plus d'un siècle, à travers (plus de) trois ordres mondiaux.** Les plus anciennes institutions de la Genève internationale remontent au XIX^e siècle. Le Comité international de la Croix-Rouge, notamment, a été fondé à Genève en 1863. La première organisation intergouvernementale formelle et permanente, l'Union télégraphique internationale, a établi ses bureaux à Berne en 1868 (avant de devenir l'Union internationale des télécommunications, basée à Genève).

Après la fin de la Première Guerre mondiale, en 1920, Genève a été choisie comme siège neutre de la première organisation multilatérale à vocation générale, la Société des Nations. C'était **la première Genève internationale.** À son apogée, la Société des Nations comptait 58 membres. Elle a toutefois été affaiblie par le refus des États-Unis

d'y adhérer et s'est finalement révélée incapable d'empêcher l'agression des puissances de l'Axe dans les années 1930. Le déclenchement de la Seconde Guerre mondiale en 1939 a montré que la Société avait failli à sa mission première.

Après la Seconde Guerre mondiale, les discussions politiques de haut niveau se sont déplacées à New York. Les États-Unis, encore irrités par les transactions des banques suisses avec l'Allemagne nazie pendant la guerre, voulaient que les Nations Unies nouvellement créées aient leur siège sur leur propre sol. Et pourtant, **la deuxième Genève internationale** a bel et bien vu le jour. Le Palais des Nations était prêt, l'Organisation internationale du Travail et le Comité international de la Croix-Rouge n'étaient jamais partis, et les rouages techniques de l'ordre multilatéral d'après-guerre se sont cristallisés autour d'eux : l'Office européen des Nations Unies, le GATT, l'Organisation mondiale de la Santé, les Conventions de Genève révisées, le CERN. La guerre froide a ensuite progressivement fait de la neutralité suisse non plus un handicap, mais un atout : Genève était l'une des rares tables autour desquelles l'Est et l'Ouest acceptaient de s'asseoir, du sommet de 1955 à la rencontre Reagan-Gorbatchev en 1985.

À la fin de la guerre froide, Genève s'est de nouveau réinventée pour devenir **la troisième Genève internationale**. La mondialisation économique a atteint de nouveaux sommets aux deux extrémités du Léman, avec l'ascension irrésistible du Forum économique mondial et la création de l'Organisation mondiale du commerce. Par effet d'entraînement, on a observé une tendance aux approches multipartites et aux partenariats public-privé plutôt qu'à des approches purement intergouvernementales. La création de Gavi, du Fonds mondial et du Forum sur la gouvernance de l'Internet relève de cette catégorie. Les discussions sur la sécurité sont également devenues moins centrées sur l'État et davantage sur l'être humain. Theodor Winkler, ancien officier du renseignement suisse, a mis à profit le passage de la Suisse à une politique de neutralité active pour faire naître une nouvelle série d'institutions genevoises consacrées à un éventail de défis sécuritaires (non traditionnels), des mines antipersonnel et du contrôle démocratique des forces armées jusqu'au terrorisme et à la gestion de crise.

Les modes de fonctionnement actuels des organisations internationales ne sont pas à l'épreuve de l'avenir. Toutefois, la concentration d'un si grand nombre d'organisations internationales à Genève continue de produire des effets d'agglomération. De même, la stabilité politique et la neutralité permanente de la Suisse restent un atout difficile à reproduire. Genève est donc un point de cristallisation logique pour des adaptations et de nouvelles institutions susceptibles d'amorcer un nouveau cycle de renouvellement, **la quatrième Genève internationale**, afin de répondre aux enjeux de coordination internationale à l'ère de l'IA.

c) Pourquoi maintenant ?

L'IA est à un tournant : 2026 est l'année où le potentiel d'usage malveillant et de perte de contrôle de l'IA est devenu sérieux, comme l'illustrent des modèles d'IA atteignant des niveaux surhumains dans la découverte de vulnérabilités. La gouvernance de l'IA perd ses petites roues. Nous connaissons déjà un certain degré d'auto-amélioration récursive, sans savoir jusqu'où se renforceront les boucles de rétroaction. Des agents de plus en plus autonomes se répandent dans l'économie plus vite que toute technologie antérieure. Nous sommes dans les "contreforts de la singularité".

La Genève internationale numérique est à un tournant : le 20 juin 2025, le Conseil fédéral a approuvé une nouvelle stratégie pour la Genève internationale, qui met particulièrement l'accent sur l'objectif d'accueillir à Genève des événements et des organisations dans les domaines de la numérisation, des technologies émergentes et de l'environnement. Le 25 juin 2025, le Conseil fédéral suisse a décidé qu'il souhaitait accueillir le Sommet de Genève sur l'IA de 2027, dans la lignée des sommets de Bletchley, Séoul, Paris et Delhi. Dès l'automne 2026 s'ouvre la "Route vers Genève", et le Sommet de juin 2027 sera un événement clé tant pour la gouvernance internationale de l'IA que pour "combler le retard numérique de Genève".

2. L'AGI Preparedness Institute

L'Institut pour la préparation à l'intelligence artificielle générale est une organisation suisse à but non lucratif fondée à Genève en 2026. Son but, tel que défini dans ses statuts, est de contribuer à préparer la gouvernance internationale et nationale aux développements de l'IA avancée et d'autres technologies transformatrices, par :

- a) la recherche et l'analyse sur l'adaptation des institutions internationales et nationales ;
- b) la publication et la diffusion de travaux à l'intention du public, des décideurs et de la Genève internationale ;
- c) la conception et l'expérimentation de nouveaux dispositifs et mécanismes de gouvernance ;
- d) le soutien aux processus de préparation menés par des institutions publiques ;
- e) l'organisation d'événements, de dialogues et de collaborations avec des institutions académiques, internationales et de la société civile ainsi qu'avec des acteurs du secteur privé.

2.1 Profil du fondateur

L'Institut est fondé et dirigé par Kevin Kohler. En 2019, il a été le premier auteur d'une proposition de "GIEC pour l'IA". En 2024-25, en tant que chercheur au Simon Institute for Longterm Governance, il a soutenu, au titre de la société civile, les consultations menées dans le cadre du processus intergouvernemental de l'Assemblée générale des Nations Unies sur le Groupe scientifique international indépendant sur l'IA et le Dialogue

mondial sur la gouvernance de l'IA. D'avril à juillet 2026, en tant que consultant individuel auprès de l'Université des Nations Unies, il a apporté un appui rédactionnel et éditorial au rapport préliminaire du Groupe. L'AGI Preparedness Institute est indépendant du Simon Institute et du Groupe de l'ONU sur l'IA, et ses positions n'engagent que lui.

Cette boucle de sept ans, "repérer tôt l'institution dont l'avenir aura besoin, s'engager dans les négociations pour la concevoir, soutenir sa mise en œuvre", est précisément celle que cet Institut a vocation à parcourir de nouveau, mais plus vite, pour l'ère de l'IAG.



2019 · Idéation — foraus : "GIEC pour l'IA"



2024-25 · Conception institutionnelle — Simon Institute : options de conception



2026 · Mise en œuvre — Groupe de l'ONU : premier rapport

Avant de fonder l'Institut, il a travaillé au Center for Security Studies de l'ETH Zurich sur la prospective, l'évaluation des risques et la gestion de crise pour l'administration fédérale suisse ; au Forum économique mondial, où il a coécrit le Global Risks Report ; et au Simon Institute for Longterm Governance, où il a dirigé la recherche et l'engagement multilatéral sur le volet de l'ONU consacré à la gouvernance de l'IA. Il a publié plus de 50 essais sur les futurs de l'IAG sur son blog "Machinocene".

2.2 Premières activités

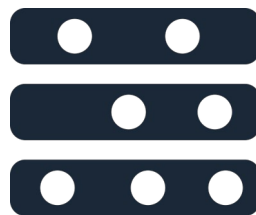
Entre septembre 2026 et le Sommet de Genève sur l'IA de juin 2027, l'Institut prévoit de travailler dans les domaines suivants :

- **Préparation de la Suisse à l'IAG.** La Suisse est-elle prête pour l'IA avancée ? Un rapport trilingue (anglais, allemand, français) fondé sur des entretiens menés au sein de l'administration fédérale, des milieux scientifiques et de l'industrie, publié en libre accès, ainsi que des contributions sur la route vers le Sommet de Genève sur l'IA de 2027.
- **Préparation du système multilatéral à l'IAG.** Publications et engagement sur la préparation du système multilatéral à l'IA avancée, notamment des contributions au Dialogue mondial sur la gouvernance de l'IA et au Sommet de Genève sur l'IA, des travaux calés sur la transition vers un nouveau Secrétaire général de l'ONU, et l'expérimentation d'outils neutres, natifs de l'IA, qui facilitent la navigation dans la documentation publique des délibérations multilatérales.
- **Idées prospectives en matière d'institutions et de processus.** Élaboration de modèles de gouvernance dont l'heure approche, par exemple un processus de type "Document de Montreux" explicitant la manière dont le droit international existant s'applique à l'IA avancée, ou un fonds de dotation cosmique (Cosmic Endowment Fund) pour les ressources spatiales relevant du patrimoine commun.

- **Ateliers et séances d'information pour les décideurs.** Exercices de scénarios d'IAG à huis clos pour les acteurs fédéraux suisses, et une série de séances d'information pour les délégations diplomatiques à Genève, en collaboration avec des institutions partenaires.

2.3 Pourquoi notre logo est un fromage suisse

Le logo de l'Institut est un empilement de trois barrières, chacune percée de trous qui ne s'alignent pas parfaitement. Il représente le modèle du fromage suisse de la gestion des risques, un modèle utilisé en ingénierie de la sécurité, de l'aviation à la riposte aux pandémies. Son message est simple : aucune couche de protection n'est parfaite. Chaque couche a des trous. Toutefois, si les couches sont indépendantes et que leurs trous ne s'alignent pas, les dangers et les acteurs malveillants peinent malgré tout à passer au travers. Voilà, en une image, la vision de l'Institut en matière de préparation à l'IAG : des modèles d'IA plus sûrs, des institutions préparées et des sociétés résilientes sont tous nécessaires, mais pas suffisants. Un écosystème d'IA aligné sur les lois et les valeurs humaines, corrigible et résilient ne peut être obtenu qu'en combinant la sûreté technique de l'IA et des institutions bien conçues. Et puis, pour un institut basé en Suisse, difficile de résister à un fromage suisse.



3. S'impliquer

L'Institut est ouvert aux collaborations. Si vous vous reconnaissez dans l'un des profils suivants, nous serions ravis d'avoir de vos nouvelles.

Pour collaborer

Si vous travaillez au sein d'une mission permanente, d'une organisation internationale, d'un gouvernement ou d'une institution de recherche et souhaitez co-organiser une séance d'information, demander un exercice de scénarios ou explorer des recherches conjointes sur les questions évoquées ci-dessus, écrivez-nous.

Pour développer des outils

Si vous souhaitez développer des outils natifs de l'IA pour de véritables processus de gouvernance, ou si vous travaillez au sein d'une institution et faites face à un goulet d'étranglement concret qu'un tel outil pourrait lever, sachez que l'Institut entend explorer des projets à cette intersection. Les manifestations d'intérêt sont les bienvenues.

Pour contribuer

L'Institut met sur pied un cercle consultatif et accueille également des contributions externes. Nous tenons à la rigueur et à la pertinence pour les processus en cours. Les auteurs en début de carrière qui ont quelque chose de substantiel à dire sont expressément les bienvenus. Proposez-nous deux paragraphes avant de rédiger le texte.

Pour faire un don

L'Institut est une association suisse à but non lucratif basée à Genève. Ses travaux sont financés par des subventions philanthropiques et des dons, et ses ressources sont irrévocablement affectées à son but statutaire. Si vous envisagez de soutenir sa mission, nous vous communiquerons volontiers des plans détaillés et des scénarios de financement.

Contact : info@agippreparedness.org · agippreparedness.org · Genève



S'abonner à la lettre d'information de l'AGIPI



AGI PREPAREDNESS INSTITUTE



agipreparedness.org