



AGI PREPAREDNESS INSTITUTE

Das vierte internationale Genf aufbauen

Das AGI Preparedness Institute stellt sich vor



September 2026

1. Warum AGI Preparedness, warum Genf, warum jetzt?¹

Das AGI Preparedness Institute ist eine neue gemeinnützige Organisation mit Sitz in Genf, die sich dafür einsetzt, nationale und internationale Institutionen auf die Folgen fortgeschrittener KI vorzubereiten. Sein englischer Name – sinngemäss "Institut für die Vorbereitung auf künstliche allgemeine Intelligenz" – spiegelt wider, dass es im internationalen Genf und mit Partnern weltweit arbeitet. Dieses kurze Einführungsdokument erläutert die Überlegungen hinter dem Institut, was es in seinem ersten Jahr vorhat und wie Sie sich beteiligen können.

a) Warum AGI Preparedness?

AGI Preparedness – die Vorbereitung auf künstliche allgemeine Intelligenz – bedeutet, die Fähigkeit von Institutionen zu stärken, die zunehmenden gesellschaftlichen, wirtschaftlichen und geopolitischen Folgen fortgeschrittener KI zu bewältigen. Dazu gehört, das Risiko von Kontrollverlust und katastrophalem Missbrauch von KI zu verringern, menschliches Gedeihen in einer Post-Arbeits-Wirtschaft zu ermöglichen und die menschliche Freiheit in einem Zeitalter der Superpersuasion und Superüberwachung zu bewahren.

Am konkretesten existiert AGI Preparedness heute in Form von Richtlinien zur verantwortungsvollen Skalierung (Responsible Scaling Policies) von Unternehmen. Dabei handelt es sich um freiwillige Verpflichtungen von KI-Unternehmen nach einer Wenn-dann-Logik: Mit zunehmendem Niveau der KI-Fähigkeiten steigen die Anforderungen an Massnahmen zur Risikominderung, etwa an eine verstärkte Cybersicherheit im Unternehmen. Das ist gut. Unternehmerische Sozialverantwortung allein reicht jedoch nicht aus:

- Die Verpflichtungen von Unternehmen zur verantwortungsvollen Skalierung sind freiwillig, und manche ihrer Aspekte wären spieltheoretisch glaubwürdiger, wenn sie überprüft und durchgesetzt werden könnten.
- Viele KI-Schäden entstehen nicht durch ein einzelnes KI-Modell, sondern durch die kumulative Nutzung vieler Systeme in der gesamten Gesellschaft.
- Bei Bedrohungen, die lediglich ein bestimmtes absolutes KI-Fähigkeitsniveau erfordern (im Gegensatz zu relativer Überlegenheit gegenüber einer anderen KI), wirken starke Kräfte zugunsten einer offenen Verbreitung, unabhängig davon, was ein einzelner Akteur tut. Das erfordert Resilienz und Notfallpläne.
- Bestimmte Risiken lassen sich wirksamer vor- oder nachgelagert zum Modell selbst steuern. KI-Modellunternehmen können weder Social-Media-Plattformen noch Internetdiensteanbieter oder bereichsspezifische Engpässe regulieren, etwa Anbieter von DNA-Synthese, die digitale Dateien in biologisches Material umwandeln.

¹ **Maschinelle Übersetzung:** Dieses Dokument wurde maschinell übersetzt. Massgebend ist die englische Originalfassung: agipreparedness.org/AGIPI_Introduction.pdf

Folglich müssen verschiedenste Akteure – von nationalen Regierungen über öffentlich-private Partnerschaften bis hin zu internationalen Institutionen – ihre Anstrengungen zur gesellschaftlichen Vorbereitung auf fortgeschrittene KI dringend verstärken. Das Institut verwendet den Begriff Artificial General Intelligence (AGI, künstliche allgemeine Intelligenz) ausdrücklich in seinem Namen, weil AGI ein zukunftsgerichteter Begriff ist. KI lässt sich nicht angemessen verstehen oder steuern, wenn man nur den heutigen Stand der Technik betrachtet. Die gesamte Rechenleistung des Bestands an KI-Chips und die Effizienz dieser Rechenleistung wachsen exponentiell. Intelligenz wird pro Jahr um rund das 40-Fache günstiger! KI-Modelle haben sich von 5% bei Mathematikaufgaben auf Gymnasialniveau im Jahr 2021 dahin entwickelt, offene Forschungsprobleme der Mathematik zu lösen, darunter im September 2026 erstmals eines der Millennium-Prize-Probleme. Von 17.5% auf Cybench im Jahr 2024 dahin, 2026 eine beispiellose Zahl realer Software-Schwachstellen zu finden. Wer zukunftsicher sein will, muss mit der Möglichkeit eines anhaltend schnellen KI-Fortschritts rechnen.

Ebenso wichtig ist es, klarzustellen, was das Institut mit AGI Preparedness nicht meint. Es geht nicht um die Vorbereitung auf einen einzigen "AGI-Tag", ein einmaliges Ereignis, nach dem die Menschheit entweder alle Probleme gelöst hat oder ausgestorben ist. Mein mentales Modell ist, dass der Übergang zu einem "Zeitalter der AGI", in dem Maschinen über die folgenreichsten Entscheidungen bestimmen werden, eher einer "komprimierten industriellen Revolution" gleicht: einer Phase anhaltenden, sich beschleunigenden Wandels statt einer einzelnen Schwelle. Ziel der Vorbereitung ist es, diesen Übergang so zu gestalten, dass die Menschheit gedeiht und dauerhaft echte Handlungsfähigkeit behält.

b) Warum Genf?

Ökosysteme aus Institutionen und Menschen können erstaunlich beständig sein. Sie schaffen spezialisierte Talent-Pipelines und implizites Wissen darüber, wie Institutionen tatsächlich funktionieren – Wissen, das sich nur lokal verbreitet. Viele haben versucht, das nächste Silicon Valley aufzubauen, und doch sind diese Bemühungen weitgehend gescheitert. Trotz all seiner Herausforderungen und Schwächen ist SF nach wie vor unangefochten die KI-Hauptstadt der Welt.

In ähnlicher Weise ist **Genf seit mehr als einem Jahrhundert und über (mehr als) drei Weltordnungen hinweg ein beständiges Ökosystem internationaler Governance**. Die ältesten Institutionen des internationalen Genf reichen ins 19. Jahrhundert zurück. Allen voran wurde das Internationale Komitee vom Roten Kreuz 1863 in Genf gegründet. Die erste formelle und ständige zwischenstaatliche Organisation, der Internationale Telegraphenverein, richtete 1868 seine Büros in Bern ein (und wandelte sich schliesslich zur Internationalen Fernmeldeunion mit Sitz in Genf).

Nach dem Ende des Ersten Weltkriegs wurde Genf 1920 als neutraler Sitz der ersten multilateralen Organisation mit allgemeinem Mandat ausgewählt: des Völkerbunds. Dies

war **das erste internationale Genf**. Auf seinem Höhepunkt zählte der Völkerbund 58 Mitglieder. Er wurde jedoch durch die Weigerung der USA, ihm beizutreten, geschwächt und erwies sich letztlich als unfähig, die Aggression der Achsenmächte in den 1930er-Jahren zu verhindern. Der Beginn des Zweiten Weltkriegs 1939 zeigte, dass der Völkerbund seinen Hauptzweck verfehlt hatte.

Nach dem Zweiten Weltkrieg verlagerten sich die hochrangigen politischen Gespräche nach New York. Die USA, frisch verärgert über die Kriegsgeschäfte Schweizer Banken mit Nazideutschland, wollten den Hauptsitz der neu gegründeten Vereinten Nationen auf eigenem Boden. Und doch entstand **das zweite internationale Genf**. Das Palais des Nations stand bereit, die Internationale Arbeitsorganisation und das Internationale Komitee vom Roten Kreuz waren nie weggezogen, und um sie herum verdichtete sich die technische Maschinerie der multilateralen Nachkriegsordnung: das Europabüro der UNO, das GATT, die Weltgesundheitsorganisation, die revidierten Genfer Konventionen, das CERN. Der Kalte Krieg machte die Schweizer Neutralität dann zunehmend von einer Belastung zu einem Trumpf: Genf war einer der wenigen Verhandlungstische, an denen Ost und West gemeinsam Platz nahmen – vom Gipfel 1955 bis zum Treffen zwischen Reagan und Gorbatschow 1985.

Als der Kalte Krieg endete, erfand sich Genf abermals neu: als **das dritte internationale Genf**. Die wirtschaftliche Globalisierung erreichte an beiden Enden des Genfersees neue Höhen, mit dem scheinbar unaufhaltsamen Aufstieg des Weltwirtschaftsforums und der neu gegründeten Welthandelsorganisation. Als Folgeeffekt zeichnete sich ein Trend zu Multistakeholder-Ansätzen und öffentlich-privaten Partnerschaften statt rein zwischenstaatlicher Ansätze ab. Die Gründung von Gavi, des Globalen Fonds und des Internet Governance Forum gehört in diese Kategorie. Auch die Sicherheitsdiskussionen wurden weniger staatszentriert und stärker auf den Menschen ausgerichtet. Theodor Winkler, ein ehemaliger Schweizer Nachrichtendienstoffizier, nutzte den Wechsel der Schweiz zu einer aktiven Neutralitätspolitik, um eine Reihe neuer Genfer Institutionen ins Leben zu rufen, die sich mit einer Vielzahl (nicht-traditioneller) Sicherheitsherausforderungen befassen sollten – von Antipersonenminen und der demokratischen Kontrolle der Streitkräfte bis hin zu Terrorismus und Krisenmanagement.

Die Arbeitsweise der heutigen internationalen Organisationen ist nicht zukunftssicher. Dennoch erzeugt die Ansiedlung so vieler internationaler Organisationen am selben Ort in Genf weiterhin Agglomerationseffekte. Ebenso bleiben die politische Stabilität und die dauernde Neutralität der Schweiz ein schwer zu kopierender Vorteil. Damit ist Genf ein logischer Kristallisationspunkt für Anpassungen und neue Institutionen, die eine weitere Runde der Selbsterneuerung einleiten können – **das vierte internationale Genf**, um Fragen der internationalen Koordination im Zeitalter der AGI anzugehen.

c) Warum jetzt?

KI steht an einem Wendepunkt: 2026 ist das Jahr, in dem das Missbrauchs- und Kontrollverlustpotenzial von KI ernst geworden ist, wie KI-Modelle zeigen, die bei der Entdeckung von Schwachstellen übermenschliche Leistungen erreichen. Die Stützräder der KI-Governance werden abmontiert. Wir haben bereits ein gewisses Mass an rekursiver Selbstverbesserung, ohne zu wissen, wie stark die Rückkopplungsschleifen werden. Zunehmend autonome Agenten verbreiten sich schneller in der Wirtschaft als jede frühere Technologie. Wir stehen "am Fuss der Singularität".

Das digitale internationale Genf steht an einem Wendepunkt: Am 20. Juni 2025 verabschiedete der Bundesrat eine neue Strategie für das internationale Genf, die insbesondere das Ziel betont, Veranstaltungen und Organisationen in den Bereichen Digitalisierung, neue Technologien und Umwelt nach Genf zu holen. Am 25. Juni 2025 beschloss der Bundesrat, den KI-Gipfel 2027 in Genf (Geneva AI Summit) auszurichten, in der Reihe der Gipfel von Bletchley, Seoul, Paris und Delhi. Ab Herbst 2026 beginnt die "Road to Geneva", und der Gipfel im Juni 2027 wird ein Schlüsselereignis sowohl für die internationale KI-Governance als auch für das "Schliessen der digitalen Genf-Lücke" sein.

2. Das AGI Preparedness Institute

Das AGI Preparedness Institute ist eine 2026 in Genf gegründete gemeinnützige Organisation nach Schweizer Recht. Gemäss seinen Statuten besteht sein Zweck darin, einen Beitrag zur Vorbereitung der internationalen und nationalen Governance auf Entwicklungen im Bereich fortgeschrittener KI und anderer transformativer Technologien zu leisten, und zwar durch:

- a) Forschung und Analyse zur Anpassung internationaler und nationaler Institutionen;
- b) Veröffentlichung und Verbreitung von Arbeiten für die Öffentlichkeit, Entscheidungstragende und das internationale Genf;
- c) die Gestaltung neuer Governance-Arrangements und -Mechanismen sowie das Experimentieren damit;
- d) die Unterstützung von Vorbereitungsprozessen öffentlicher Institutionen;
- e) die Organisation von Veranstaltungen, Dialogen und Kooperationen mit akademischen, internationalen und zivilgesellschaftlichen Institutionen sowie mit Akteuren des Privatsektors.

2.1 Profil des Gründers

Das Institut wurde von Kevin Kohler gegründet und wird von ihm geleitet. 2019 war er Erstautor eines Vorschlags für ein "IPCC für KI". 2024–25 unterstützte er als Forscher am Simon Institute for Longterm Governance im Rahmen der Zivilgesellschaft die Konsultationen während des zwischenstaatlichen Prozesses der UNO-

Generalversammlung zum Unabhängigen internationalen wissenschaftlichen Gremium für künstliche Intelligenz und zum Globalen Dialog über die Lenkung künstlicher Intelligenz. Von April bis Juli 2026 unterstützte er als Einzelberater der Universität der Vereinten Nationen das Verfassen und Redigieren des vorläufigen Berichts des Gremiums. Das AGI Preparedness Institute ist unabhängig vom Simon Institute und vom KI-Gremium der UNO; seine Ansichten sind seine eigenen.

Dieser siebenjährige Zyklus – “die benötigte künftige Institution früh erkennen, sich an den Verhandlungen zu ihrer Gestaltung beteiligen, ihre Umsetzung unterstützen” – ist genau der Zyklus, den dieses Institut erneut durchlaufen soll, nur schneller und für das Zeitalter der AGI.



2019 · **Ideenfindung** — foraus: “IPCC für KI”



2024-25 · **Institutionelle Gestaltung** — Simon Institute: Gestaltungsoptionen



2026 · **Umsetzung** — UNO-Gremium: erster Bericht

Vor der Gründung des Instituts arbeitete er am Center for Security Studies der ETH Zürich zu Foresight, Risikobewertung und Krisenmanagement für die Bundesverwaltung, beim Weltwirtschaftsforum, wo er Mitautor des Global Risks Report war, sowie am Simon Institute for Longterm Governance, wo er die Forschung und das multilaterale Engagement zum KI-Governance-Strang der UNO leitete. Auf seinem Blog “Machinocene” hat er mehr als 50 Essays über AGI-Zukünfte veröffentlicht.

2.2 Erste Aktivitäten

Zwischen September 2026 und dem KI-Gipfel in Genf im Juni 2027 plant das Institut, in folgenden Bereichen zu arbeiten:

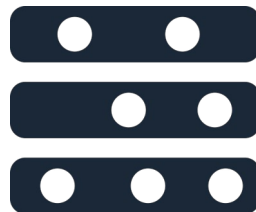
- **AGI-Bereitschaft der Schweiz.** Ist die Schweiz bereit für fortgeschrittene KI? Ein dreisprachiger Bericht (Englisch, Deutsch, Französisch) auf der Grundlage von Interviews in der Bundesverwaltung, der Wissenschaft und der Industrie, frei zugänglich (Open Access) veröffentlicht – sowie Beiträge auf dem Weg zum KI-Gipfel 2027 in Genf.
- **Multilaterale AGI-Bereitschaft.** Publikationen und Engagement zur Vorbereitung des multilateralen Systems auf fortgeschrittene KI – darunter Beiträge zum Globalen Dialog über die Lenkung künstlicher Intelligenz und zum KI-Gipfel in Genf, Arbeiten, die auf den Übergang zu einer neuen UNO-Generalsekretärin oder einem neuen UNO-Generalsekretär abgestimmt sind, sowie Experimente mit neutralen, KI-nativen Werkzeugen, die die öffentliche Dokumentation multilateraler Beratungen leichter zugänglich machen.
- **Zukunftsgerichtete Ideen für Institutionen und Prozesse.** Entwicklung von Governance-Modellen, deren Zeit kommt – zum Beispiel ein Prozess für ein “Montreux-Dokument”, das darlegt, wie bestehendes Völkerrecht auf

fortgeschrittene KI anwendbar ist, oder ein Cosmic Endowment Fund für Weltraumressourcen als gemeinsames Erbe der Menschheit.

- **Workshops und Briefings für politische Entscheidungstragende.** Nichtöffentliche AGI-Szenarioübungen für Anspruchsgruppen des Bundes sowie eine Briefing-Reihe für diplomatische Delegationen in Genf, gemeinsam mit Partnerinstitutionen.

2.3 Warum unser Logo ein Schweizer Käse ist

Das Logo des Instituts ist ein Stapel aus drei Barrieren, jede von Löchern durchbrochen, die nicht ganz übereinanderliegen. Es steht für das Schweizer-Käse-Modell des Risikomanagements, ein Modell, das in der Sicherheitstechnik von der Luftfahrt bis zur Pandemiebekämpfung verwendet wird. Seine Botschaft ist einfach: Keine einzelne Schutzschicht ist perfekt. Jede Schicht hat Löcher. Wenn die Schichten jedoch unabhängig voneinander sind und ihre Löcher nicht übereinanderliegen, kommen Gefahren und Bedrohungsakteure trotzdem nur schwer hindurch. Das ist das Verständnis des Instituts von AGI Preparedness in einem Bild: Sicherere KI-Modelle, vorbereitete Institutionen und resiliente Gesellschaften sind allesamt notwendig, aber nicht hinreichend. Ein KI-Ökosystem, das mit menschlichen Gesetzen und Werten im Einklang steht, korrigierbar und resilient ist, lässt sich nur durch eine Kombination aus technischer KI-Sicherheit und gut gestalteten Institutionen erreichen. Und einem Schweizer Käse konnte ein Institut mit Sitz in der Schweiz ohnehin kaum widerstehen.



3. Mitwirken

Das Institut ist offen für Zusammenarbeit. Wenn eine der folgenden Beschreibungen auf Sie zutrifft, freuen wir uns, von Ihnen zu hören.

Als Partnerin oder Partner

Wenn Sie für eine ständige Vertretung, eine internationale Organisation, eine Regierung oder eine Forschungseinrichtung tätig sind und gemeinsam mit uns ein Briefing ausrichten, eine Szenarioübung anfragen oder gemeinsame Forschung zu den oben genannten Fragen ausloten möchten, schreiben Sie uns.

Als Entwicklerin oder Entwickler

Wenn Sie KI-native Werkzeuge für reale Governance-Prozesse entwickeln möchten oder innerhalb einer Institution tätig sind und einen konkreten Engpass haben, den ein solches Werkzeug beheben könnte: Das Institut beabsichtigt, Projekte an dieser Schnittstelle auszuloten. Interessenbekundungen sind willkommen.

Als Autorin oder Autor

Das Institut baut einen Beratungskreis auf und freut sich auch über Gastbeiträge. Uns sind Gründlichkeit und Relevanz für laufende Prozesse wichtig. Autorinnen und Autoren am Anfang ihrer Laufbahn, die wirklich etwas zu sagen haben, sind ausdrücklich willkommen. Schicken Sie uns zwei Absätze als Pitch, bevor Sie den Beitrag schreiben.

Als Spenderin oder Spender

Das Institut ist ein gemeinnütziger Verein nach Schweizer Recht mit Sitz in Genf. Seine Arbeit wird durch philanthropische Zuwendungen und Spenden finanziert, und seine Mittel sind unwiderruflich seinem statutarischen Zweck gewidmet. Wenn Sie erwägen, die Mission zu unterstützen, teilen wir gerne detaillierte Pläne und Finanzierungsszenarien mit Ihnen.

Kontakt: info@agippreparedness.org · agippreparedness.org · Genf



AGIPI-Newsletter abonnieren



AGI PREPAREDNESS INSTITUTE



agipreparedness.org