Harnessing AI Risk Initiative

Executive Summary

The *Harnessing AI Risk Initiative* is an ongoing effort aimed at aggregating a critical mass of globally diverse states to jump-start and design an **open, expert and participatory constituent process for the creation of new global intergovernmental organizations for AI and digital communications,** that are suitable to reliably manage their immense risks in terms of human safety and concentration of power and wealth, and realize their potential to usher us in an era of unimagined prosperity, safety and well-being.

NOTE: The text below is a **April 18th 2024** copy of the web pages of the Harnessing AI Risk Initiative, its 1st Harnessing AI Risk Summit and the Opportunities pages - with the addition of 4 chapters after the Initiative's page, and a chapter on the *Global Public Benefit AI Lab*.

Table of Contents

Executive Summary

(the text below is the same text as the Initiative's web page on April 11th, 2024)

Introduction two-pager

The *Harnessing AI Risk Initiative* is an ongoing effort aimed at aggregating a critical mass of globally diverse states to jump-start and design an **open, expert and participatory constituent process** for the creation of a new **global intergovernmental treaty-organization for AI and digital communications** that is suitable to reliably manage their immense risks in terms of human safety and concentration of power and wealth, and realize their potential to usher us in an era of unimagined prosperity, safety and well-being.

The Initiative is calling on a few and then a critical mass of globally-diverse states to join in summits in Geneva to agree on the *Scope and Rules for the Election* of an ***Open Transnational Constituent Assembly for AI and Digital Communications.*** Given the inherently global nature of AI's primary threats and opportunities, the mandate of such an *Assembly* will need to include the following:

- Setting global AI safety, security and privacy standards
- Enforcing global bans for unsafe AI development and use
- Developing world-leading or co-leading safe AI capabilities via a public-private $15+ billion *Global Public Benefit AI Lab* and supply chain
- Developing globally-trusted *governance-support* systems

The design of such an *Assembly* will aim to maximize **expertise, timeliness and agility**, on the one hand, and **participation, democratic process, neutrality and inclusivity**, on the other, to maximize the chances that the resulting organization will be sufficiently trustworthy and widely trusted to:

- Encourage broad compliance with future bans and oversight
- Enhance safety through diversity and transparency in setting standards
- Ensure a fair and safe distribution of power and wealth
- Mitigate destructive inter-state competition and global military instability

A key milestone will be the 1st Harnessing AI Risk Summit this November 2024 in Geneva, preceded by a *Pre-Summit Virtual Conference* on June 12th among NGOs and experts.

A Better Treaty-Making Method

Unfortunately, the prevailing treaty-making methods used to build new intergovernmental organizations have been broken for decades, as demonstrated by those for climate change and nuclear weapons.

The Initiative will, therefore, largely replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

Voting weight in the Assembly will be **adjusted primarily according to population size and GDP** - also in consideration of the current huge asymmetry in AI capabilities and world power and the fact that 3 billion persons remain and/or illiterate. The emphasis on GDP will be bindingly reduced, in a few years, as the organization will have ensured nearly all are literate and connected. States and superpowers that will join early will have substantial but temporary economic and voting-power advantages.

Strategic Positioning

The Initiative seeks to fill the **wide gaps in global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs and firms - including the US, China, the EU, the UN and OpenAI's public-private "trillion AI plan" - and become the platform for their convergence.

The Initiative aims to become the key enabler of the call by the ***UN Secretary-General*** *for an "IAEA for AI."* It aims to build a treaty-making *vehicle* that has the global legitimacy and representativity that is needed, and his office, agencies and boards are lacking - in line with his clarification that "*only member states can create it, not the Secretariat of the United Nations*." The Initiative will eventually constitute a *Caucus within the UN General Assembly and later seek approval by the UN General Assembly to become a part of the UN system* while retaining full governance autonomy.

As in 1946, when the US and Russia proposed a new independent UN agency to manage all nuclear weapons stockpiles and weapons and energy research via their Baruch and Gromyko Plans but failed to agree, **we now have a second chance with AI**. We can harness AI's risk to turn it into an unimagined blessing for humanity and set a governance model for other dangerous technologies and global challenges.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its upcoming *Summit* in November or *Pre-Summit Virtual Conference* on June 12th.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 missions' AI and digital domain experts, and we are engaging 3 more. Together, those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we have been in extended talks with 3 of the 5 AI Labs about their interest in participating in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participants to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Summit this November in Geneva.

Preliminary Designs and Scope of the new IGO

The Initiative is also advancing - in unique levels of detail and comprehensiveness, and with the support of dozens of advisors and experts - a *proof-of-concept* proposal for the **scope**, **functions and character of such a new intergovernmental organization** that match the scale and nature of the challenge.

We group the required functions in three agencies of a single IGO, subject to a **federal, neutral, participatory, democratic, resilient, transparent and decentralized governance structure** with effective checks and balances:

- (1) An **AI Safety Agency** will set global safety standards and enforce a ban on all development, training, deployment and research of dangerous AI worldwide to sufficiently mitigate the risk of loss of control or severe abuse by irresponsible or malicious state or non-state entities.
- (2) A **Global Public Benefit AI Lab** will be a $15+ billion, open, partly decentralized, democratically governed joint venture of states and suitable tech firms aimed at

achieving and sustaining solid global leadership or co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures.
    - It will accrue member states' capabilities and resources and distribute dividends and control to member states and directly to their citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.
    - It will be primarily funded via *project finance*, buttressed by pre-licensing and pre-commercial procurement from participating states and firms.
    - It will seek to achieve and sustain a resilient "mutual dependency" in its wider AI supply chain - vis-a-vis AI superpowers and other future consortia - through joint investments, diplomacy, trade relations and strategic industrial assets of participant states.
- (3) An **IT Security Agency** will develop and certify radically more trustworthy and widely-trusted AI governance-support systems, particularly for confidential and diplomatic communications, for control subsystems for frontier AIs and other critical societal infrastructure, such as social media.

Far from being a fixed blueprint, such a proposal aims to fill a glaring gap in the availability of detailed and comprehensive proposals. It aims to stimulate the production of other similarly comprehensive proposals to foster concrete, cogent, transparent, efficient, and timely negotiations among nations leading up to such Assembly and eventually arrive soon at *single-text procedure* negotiations based on majority and supermajority rule, rather than unanimity.

Learning from History's Greatest Treaty-Making Success

Nine years after the *U.S. Articles of Confederation* were enacted in 1781, many U.S. states realised it was far from enough to safeguard both their economy and their security.

Hence, two of them convened three others in the Annapolis Convention in 1786, and decided to design and convene a U.S. Constitutional Convention for 1787, to build a true federation.

There, state delegations agreed by simple majority on a U.S. Constitution bound to come into force if 9 out of 13 states legislatures approved it. In hindsight, it was an astounding success, except only 1 out of 8 adults had voting rights.

A similar process, and for the same reasons, can and should be replicated at the global level for AI—a history-defining technology with immense implications for the economy, safety, security and human nature.

Once we succeed in gathering 7 or more globally diverse states, it will be relatively easy through then to attract dozens more to have a successful "global Annapolis Convention for AI".

Opportunities

Find below detailed opportunities to join, support or partner with the Harnessing AI Risk Initiative and/or its 1st Harnessing AI Risk Summit this November 2024, in Geneva:

- Opportunities for **States**
- Opportunities for **Donors**
- Opportunities for **NGOs and Experts**
- Opportunities for **Funders and Investors** of the Lab
- Opportunities for **Leading AI Labs**
- Opportunities for **Regional Intergovernmental Organizations**

About Us

The *Trustless Computing Association* is a Geneva-based non-profit with a mission to promote safe, secure and democratic IT and AI by fostering the creation of new intergovernmental organizations, socio-technical security paradigms and technologies. It does so via institution-building initiatives supported by research initiatives, publications, the TRUSTLESS.AI spin-in (closed Sept 2023), and via 11 editions of the Free and Safe in Cyberspace on 3 continents.

Until March 2013, its activities were centered on the Trustless Computing Certification Body (TCCB) and Seevik Net Initiative. Since then, our focus has moved on to the *Harnessing AI Risk Initiative aimed* at the creation of a new IGO with three agencies to manage AI, including the TCCB and its Summit series. The association is supported by 32 world-class advisors and over 25 partners. See About Us and Team and Advisors page for more.

Full Information on the Initiative in a single PDF

A 63-page **Executive Summary of the Harnessing AI Risk Initiative and Summit PDF**. (live updated) A copy of the web pages of the *Harnessing AI Risk Initiative*, its *1st Summit*, Pre-Summit, and the Opportunities pages for states, IGOs, donors, NGOs, AI labs and investors in the Lab. Includes also a 6-page chapter on the *Global Public Benefit AI Lab*.

Other Publications, Articles and Posts

- *Can a global version of the 1786 Annapolis Convention lead to the governance we need for AI?. (*A March 2024, 1200-word blog post)
- *How a public-private consortium could lead to democratic global AI governance.* (A March 2024, 900-word opinion piece the president of the Trustless Computing Association published last March 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis global AI supply chains, OpenAI's "$7 trillion AI plan", and the pursuit of an effective, democratic and safe global governance of AI).
- A 33-page Harnessing AI Risk Proposal v.3 PDF, (published January 2024, on *ResearchGate*). It details the Initiative, its rationale, the design of the constituent processes, and the preliminary designs of the IGO and its agencies. It sets an initial framework for the Initiative's co-design with advisors, partners and Summit participants. (*Harnessing AI Risk Proposal v.2* (Oct 2023, 6500-word paper, published on ResearchGate pdf) and *Harnessing AI Risk Proposal v.1* (June 2023, published as Linkedin and blog post)
- A 14-page Grant Proposal and Roadmap 2024-2027 of the Initiative PDF. (March 10th, 2024). Includes 1-page summary.
- A February 2024, 8-page Case for family offices to support and invest in a Global Public Benefit AI Lab and an International AI Safety Agency.
- A December 2023, 700-word blog post, The AI Act and Beyond: EU's Ambitions and Obstacles in the AI Race. It frames our Initiative vis-a-vis EU AI Act and EU AI capacity-building initiatives.
- An October 2023, 3000-word blog post, Towards an Open Transnational Constituent Assembly for AI and Digital Communications.
- For further details about the foreseen ***IT Security Agency*****,** in addition to *Harnessing AI Risk Proposal* (v.3) above, see our Trustless Computing Certification Body and Seevik Net Initiative (1-pager + 45-pager pdf) and details of our traction so far with over 13 nation-states (1-pager + 32-pager pdf).

AI as our Greatest Risk

The alarm has sounded for the immense risks posed by AI, along with its great opportunities.

Since March 2023, when hundreds of AI **scientists,** including two of the top three, stated that "*Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war*", awareness of AI safety risk has been skyrocketing. Twenty-eight **states**, accruing to 80% of the world population, acknowledged in the Bletchley Declaration such safety risks, including "loss of control." Over 55% of **citizens** surveyed in 12 developed countries were "fairly" or "very" worried about "loss of control over AI." At an invitation-only CEO Summit at Yale last June, 42% of **CEOs** surveyed said they believed AI has the potential to "destroy humanity within the next five to 10 years."

The risks of AI leading to **extreme unaccountable concentration of power and wealth**—including via misinformation, surveillance, manipulation, oligopolies and biases—are just as important and urgent, and awareness of this is just as widespread among states and citizens.

Frontier AI capabilities are expected to keep **expanding five to tenfold annually**. And that's based on growth in investments and computing power alone, without accounting for AI's increasing **ability to self-improve** and multiply the productivity of its developers. A break-neck **AI arms race** among nations and firms is unfolding.

Meanwhile, seven years after the Cambridge Analytica scandal and ten after the Snowden revelations, **social media and sensitive communications** are ever more vulnerable to abuse and control by unaccountable entities, stifling fair and effective dialogue, within and among nations, at a time when it is most needed.

Investments in AI and AI infrastructure are exploding. If successful, OpenAI's public-private **$7 trillion AI plan** to aggregate states, funders, chip makers and power providers will either (a) create an entrenched dominant global oligopoly under US control or else (b) possibly become the seed of a safe and democratic global governance of AI that Altman has been consistently calling for - as we argue in this article in The Yuan.

AI as Our Greatest Opportunity

If we manage to avert catastrophic risks to safety and concentration of power by creating proper global AI governance institutions, the benefits of human-controllable and

humanity-controlled AI will be astounding and largely unimaginable in terms of abundance, peace, safety and well-being**.**

The potential "AI pie," if we avoid the immense risks, is so enormous that rich states and people can get richer while the poor can be much better off. But success inevitably requires a fair distribution of the power in shaping our collective future in this Digital and AI Age.

As in 1946, when the US and Russia, with their Baruch and Gromyko Plans, proposed a new independent UN agency to manage all nuclear weapons stockpiles and weapons and energy research but failed to agree, **we now have a second chance with AI**. We can harness AI's risk to turn it into an unimagined blessing for humanity and set a governance model for other dangerous technologies and global challenges.

Why Treaty-Making for AI is Broken

Leading digital and AI **superpowers** appear locked in a reckless arms race - economic, military and geopolitical - over AI and AI chips, seemingly intent on hegemonizing it or, at best, eventually splitting its global dominance.

On their own, **nearly all states stand powerless in the face of AI**, unable to avoid its immense risks for safety, for the concentration of power and wealth, and unable to realize its astounding opportunities. Even larger states like Brazil, India and Germany, or large confederations like the EU. On their own, **nearly all nations lack the strategic autonomy**, on their own, to table more democratic constituent processes to safeguard their economy, sovereignty, and safety in such all-important domains.

Initiatives for the global governance of AI are apparently led by existing **Intergovernmental organizations and fora**, like the UN, G7, G20, the EU, Council of Europe, OECD, GPAI, and the AI Safety Summits.

Yet, these are structurally unable to lead a *democratic* or effective global constituent process for AI governance due to their lack of a mandate, lack of representativity, closed membership and/or statutory over-reliance on unanimity decision-making. Hence, their initiatives severely lack multilateralism, detail, timeliness, breadth, transparency and global inclusivity and are mostly controlled by a handful of states.

The **prevailing treaty-making models** are bound to result in **severely weak, fragile and undemocratic treaties** - as they did largely in past decades - due to their reliance on loose, undefined, unstructured processes, over-reliant on unanimity.

The real explanation is that those treaty-making initiatives are really smokescreen and distractions from the real negotiations. While intergovernmental organizations and fora engage in hopeless treaty-making processes for AI that are structurally weak, slow and undemocratic, **global governance of AI is really taking shape via competition and negotiations among superpowers** and their national security agencies - with an observer role for selected allies - as it did for all other disrupting technologies in the past.

While those superpowers managed to avoid the realization of the most catastrophic risks of nuclear and bioweapons (so far!), the risks of nuclear and bioweapons are today higher than they ever were and instituted an opaque surveillance apparatus that has undermined democracy worldwide. In addition, mitigating the **proliferation and safety risks of AI is likely to be much harder than nuclear**, and therefore a much wider global compliance, adherence, and participation to common safety rules will be necessary.

After the UK AI Safety Summit was convened to foster international cooperation on AI risks of misuse and loss of control, the **United States** and the UK each announced their own AI safety institutes instead. A month later, Guidelines for Secure AI System Development were published last November by the national security agencies of the **United States** and the UK, together with the cybersecurity standardization bodies of 16 allied states. Meanwhile, **China** announced its Global AI Governance Initiative, which calls for a "*United Nations framework to establish an international institution to govern AI*" that will "*ensure equal rights, equal opportunities, and equal rules for all countries in AI development and governance.*", but no action followed, except for some reported discussions with the US on AI.

While recent US announcements that it intends to "cooperate" with China on AI safety are very welcome news, albeit late, control of AI safety by a handful of states, as was done for nuclear after WW2, would not work for two main reasons. It would be unfair and undemocratic, leading to **extreme and unaccountable concentrations of power and wealth**. It would likely not work for safety either because the nuclear threat is higher today than it ever was and because preventing catastrophic AI proliferation will likely require a much wider global adoption and compliance.

While lacking so far in inclusion, transparency and democratic process, such US/UK initiative and US/China talks are highly welcome, given that so much of the relevant expertise accrues in their security agencies, and that several top AI experts think **catastrophic safety risks may be just years away**, including via leaked, stolen or published dangerous LLM weights.

Hence, we need a treaty-making model that reconciles global legitimacy, democracy, expertise and reckoning with the huge asymmetries of power and expertise in AI among states.

Calls by AI labs for Democratic Global Governance of AI

Several of the leading AI labs, their CEOs and top AI scientists - more keenly aware than citizens and heads of state of the immense risks for safety and concentration of power - have called clearly for strong and democratic global governance of AI and in some cases for using the model of the **global constituent assembly**.

Google DeepMind published last July a detailed "exploration" of the feasibility of creating four new IGOs for AI, including a Frontier AI Collaborative, an "*international public-private partnership*" to "*develop and distribute cutting- edge AI systems, or to ensure such technologies are accessible to a broad international coalition*". Its CEO stated in February he sees in the next few years their governance merging into a UN-like organization as we get closer to "AGI".

As mentioned above, we are literally taking **OpenAI**'s CEO Sam Altman at its word - and holding him to it! - when he **called for a global constituent assembly akin to the U.S. Constitutional Convention of 1787** to establish a federal intergovernmental organization to manage AI, in a decentralized and participatory way, according to the *subsidiarity principle*.

Far from an extemporaneous statement, it was largely confirmed in later video interviews yet pushed “down the road”. He stated that control over OpenAI and advanced AI should eventually be distributed among all citizens of the world. He stated that “*we shouldn’t trust*” OpenAI unless its board "***years down the road will not have sort of figured out how to start***” transferring its power to "all of humanity"”

He stated if humanity jointly decided that pursuing “AGI” was too dangerous, they would stop all “AGI” development ("We'd respect that"). After OpenAI’s governance crisis, he repeated that people shouldn’t trust OpenAI unless it democratizes its governance. He then repeated that all of humanity should be shaping the future of AI. On February 24th, OpenAI stated in its revised mission, “*We want the benefits of, access to, and **governance** of AGI to be widely and fairly shared*.”

Given the acceleration in AI capabilities, investment and concentration in recent months, and OpenAI’s proposal of a public-private “***$7 trillion AI supply chain plan***”**,** we believe that his pledge to transfer such power “*years down the road*” sounds more and more like an empty promise, unless they are turned very soon into precise timelines and modalities for the transfer of power to humanity. Yet, as he appropriately stated at the *World Government Summit*, “it is not up to them” to define such constituent processes, so he called on states, such as the UAE, to convene a Summit aimed at the creation of an “IAEA for AI.”

Anthropic's CEO Dario Amodei suggested (7-minutes onwards in this video) that solving the *technical half* of the AI alignment problem would be of no use unless the global *governance half* is also solved and that eventually some global body should be in charge of all advanced AIs companies. It has experimented with Collective Constitutional AI to enable (national) citizens' assemblies to determine the values and constraints of AI, a process that could be extended to world citizens and to states.

OpenAI's Chief Scientist **Ilya Sutzkever** stated, "*it will be important that AGI is somehow built as a cooperation between multiple countries*." **Yoshua Bengio** called for a multilateral network of AI labs, analyzing in fine detail the right balance of global and national authority over them. The Initiative aligns with several open calls, such as for an AI Treaty, signed by many top AI scientists, to create both an "IAEA for AI" and a "CERN for AI", as well as those by The Elders and Future of Life Institute, and by Pope Francis, though none of them calls for a democratic constituent process.

While important and encouraging, those calls have diminished in recent months and do not tackle the all-important issue of the **nature, details, participation and timing of the constituent process** to arrive at such treaties that would overall most likely promote global public interest.

The Global Public Benefit AI Lab

The Global Public Benefit AI Lab (or "Lab") will be an open, democratically-governed joint-venture of states and AI labs aimed to achieve and sustain a solid global leadership or co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures. It will accrue capabilities and resources of member states and firms, and distribute dividends and control to member states, while stimulating and safeguarding private initiative for innovation and oversight.

At A Glance

The **Global Public Benefit AI Lab** will be a $15+ billion, open, partly-decentralized, democratically-governed joint-venture of states and suitable tech firms aimed to achieve and sustain a solid global leadership or co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures.

The *Lab* is one of three agencies of a new intergovernmental organization being built by the Harnessing AI Risk Initiative, a venture to catalyze a critical mass of globally-diverse states in a global constituent processes to build a new democratic IGO and joint venture - open to all states and firms to join on equal terms - to jointly build the most capable safe AI, and reliably ban unsafe ones.

- The Lab will be an open, partly-decentralized, democratically-governed joint-venture of states and suitable tech firms aimed to achieve and sustain a solid global leadership or co-leadership in ***human-controllable* AI capability**, technical alignment research and AI safety measures.
- The Lab will accrue **capabilities and resources** of member states and private partners, and distribute dividends and control among member states and directly to their citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.
- The Lab will be **primarily funded via *project finance***, buttressed by pre-licensing and pre-commercial procurement from participating states and client firms.
- The Lab will seek to achieve and **sustain a resilient "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms, as detailed in our recent article on *The Yuan*.

Financial Viability and the Project Finance model

The Lab will generate revenue from governments, firms and citizens via licensing of enabling back-end services and IP, leasing of infrastructure, direct services, and issuance of compliance certifications.

Given that the proven scalability of capabilities, value-added and profit potential of current open source LLMs technologies - and the possibility of extensive pre-commercial procurements contracts with states could buttress its financial viability - the initial funding could follow primarily the project finance model, via **sovereign and pension funds, intergovernmental sovereign funds such as the EIB and AIB, sovereign private equity and private international finance**.

The undue influence on the governance of private funding sources will be limited via various mechanisms, including non-voting shares.

Precedents

The initiative could take inspiration from the current governance of the CERN, a joint venture for nuclear energy capability-building that was started in 1954 by EU states and only later opened to non-EU ones, with a current yearly budget of $1.2 billion. The $20 billion international consortium ITER for nuclear fusion energy is also an inspiration.

Size of Initial Funding

Since the cost of state-of-the-art LLMs "training runs" are expected to grow 500-1000% per year, and many top US AI labs have announced billion-dollar LLM training runs for next year, the Lab would need an initial endowment of **at least $15 billion** to have a solid chance of achieving its capacity and safety goals, and then financial self-sustenance in 3-4 years. If such an amount seems high, consider it would likely increase by about 5-10 times for every year this initiative is delayed.

Supply-Chain Viability and Control

Acquiring and maintaining access to the specialized AI chips needed to efficiently run leading-edge LLM training runs will be challenging given a foreseen intense increase in global demand and export controls.

This is a risk that can likely be sufficiently reduced via joint diplomatic dialogue, appealing to the open and democratic nature of the initiative, and by attracting participating states hosting

firms owning suitable AI chips designs, or possibly start pursuing its own AI chip designs, and chip manufacturing capabilities, and invest in new safer and more powerful AI software and hardware architectures, beyond large language models.

Ensuring sufficient energy sources, suitable data centers, and resilient network architecture among the member states, would require timely, speedy and coordinated action for the short term and careful planning for the long term.

Hence, the Lab will seek to achieve and **sustain a resilient "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms, as detailed in our recent article on *The Yuan*.

Talent Attraction Feasibility

Key to achieving and retaining a decisive superiority in advanced AI capability and safety - especially if or until AI superpowers and their firms have not joined - is the ability to attract and retain top AI talent and experts. Talent attraction in AI is driven by compensation, social recognition and mission alignment and would need to ensure very high security and confidentiality.

Staff will be paid at their current global market value, and their social importance will be highlighted. Member states will be mandated to support top-level recruitment and to enact laws that ensure that knowledge gained is not leaked. Staff selection and oversight procedures will exceed those of the most critical nuclear and bio-labs facilities in sophistication.

The unique mission and democratic nature of the Lab would likely have a strong chance of being perceived by most top global AI researchers, even in non-member states, as being ethically superior to others, akin to how Open AI originally, and Meta more recently, have attracted top talent to work with them, or for them, via claims of their "open-source" ethos.

Just as OpenAI attracted top talent from Deepmind due to a mission and approach perceived as superior, and top talents from OpenAI went on to create Anthropic for the same reasons, the Lab should be able to attract top talents as the next "most ethical" AI project. Substantial risks of authoritarian political shifts in some AI superpowers, as warned (1.5 min video clip) by Joshua Bengio, could entice top talents to join the Global AI Lab to avoid their work being instrumental to an authoritarian regime.

Public-Private Partnership Model

Participant AI labs would join as *innovation and go-to-market partners*, in a joint-venture or consortium controlled by the participant states.

They will contribute their skills, workforce and part of their IP in such a way as to **advance both their mission to benefit humanity, their stock valuations**, and retain their agency to innovate at the root and application level, within safety bounds:

- As *innovation partners* and IP providers, they would be compensated via revenue share, secured via long-term pre-licensing and pre-commercial procurement agreements from participating states and firms.
- As *go-to-market partners*, they would gain permanent access to the core AI/AGI capabilities, infrastructure, services and IP developed by the Lab.
    - These will near-certainly far outcompete all others in capabilities and safety, and be unique in actual and perceived trustworthiness of their safety and accountability.
    - They would maintain the freedom to innovate at both the base and application layers, and retain their ability to offer their services to states, firms and consumers, within some limits.
- Participant AI labs partnership terms will be designed so as to maximize the chances of a steady increase in their market valuation, in order to attract the participation of AI labs - such as Big Tech firms - that are governed by legal conventional US for-profit vehicles that legally mandate their CEOs to maximize shareholder value.

This setup will enable such labs to continue to innovate in capabilities and safety at the base and application layers but outside a “Wild West" race to the bottom among states and labs, advancing both mission and market valuation.

Alternatively, to a public-private model the Lab could adopt a spin-in model, such as that utilized by the **German Department of Defense** to develop highly-skilled national and foreign firms in the development of critical IT infrastructure, via its German DoD Hercules. TheThis arrangement ensures access to private sector innovation and long-term multi-national highly-democratic and globally-representative control of the digital sphere resulting from mandatorily interoperable TCCB systems, Seevik Net.

The Superintelligence Option

It is of great importance that nearly all leading US AI firms - while acknowledging the real and huge human safety risks of AGI or Superintelligence - have publicly committed to pressing on to build it, asserting each that their specific approaches will maintain human control over these systems, and/or that their emergence is unlikely to be stopped. This raises the legitimate question of whether some of these AI labs might be comfortable with a significant risk of humanity losing control over AI, or even hiddenly rooting for it.

Their rationale is hinted at in recent interviews and publications. Some of them believe that it's too hard to stop all advanced private and state entities from pursuing it, and therefore, they should try to do their best to influence its nature, if at all possible. Perhaps, they also consider it probable or plausible that an AI takeover may cause a good or great case scenario for humanity or valuable future life forms.

Calls for a global lab and governance by top US labs and NGOs

Participation could be extended to AI labs from states that may initially not be member states of the new organization, such as one or more AI superpowers.

While governments have shown reluctance, leading AI labs - which stand to lose the most from stringent or global regulations - have been the most outspoken about both the "catastrophic safety risks" and the necessity for global governance. Some have even presented highly detailed proposals, while others have made explicit calls for democratic and participatory frameworks.

Perhaps due to their substantial global lead over their competitors in other states, several US leading AI labs have called for one or more of the following: (1) enforce a global cap and ban on dangerous AI developments, (2) reduce the "race to the bottom on safety" (3) insurance of more time, resources and coordination in tackling the technical alignment problem; and (4) solve via globally democratic governance the governance alignment problem.

There is a strong awareness of the catastrophic safety risks of a global AI arms race among half of the top US AI labs and top Chinese and US AI scientists. The CEOs of Deepmind, OpenAI and Anthropic, and 2 of the three "godfathers of AI", signed with many others a letter in May 2023 stating, "*Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.*"

In a development largely overlooked by mainstream media, Sam Altman, the CEO of **OpenAI**, which developed ChatGPT, last March called for a global democratic constitutional convention similar to the 1787 U.S. Constitutional Convention. The aim would be to establish a global

federal governance system for AI based on the principle of *subsidiarity*. He repeated calls for highly participatory, federal and empowered global governance of AI in an interview, and then another. He even pledged to transform OpenAI's current governance structure in a globally democratic and participatory body, stating that OpenAI should not be trusted if its 7-person board fails to enact soon such a transfer of power.

Last November 17th, 2023, Altman was fired by the OpenAI board. After 747 of 770 of its employees requested him as CEO, the board was forced to resign. While many read the events as a sign that OpenAI and Altman are beholden by profit motives and pressures by Microsoft and investors, we believe it arguably resulted in a much-increased *de-facto* power to shape OpenAI's future governance.

In a December 7th audio interview he gave in 2023 to Trevor Noah, Altman acknowledged as much and stated those same intentions for the future governance of OpenAI, with an even more explicit call for a role of governments. Again, on December 13th, he reiterated in an interview (1-minute video clip) how people should not trust OpenAI unless its governance does not democratize.

Similarly, Dario Amodei, the CEO of **Anthropic**, one of the top 5 AI firms in the world, loudly warned of significant and near-term catastrophic risks, and called for strong democratic global governance of AI, and suggested that their planned controlling non-profit steering board should be absorbed or replaced at some point by a global democratic body.

OpenAI's Chief Scientist **Ilya Sutskever** stated in minutes 9.51-10.43 of this documentary: "I*f you have an arms race dynamic between multiple kings trying to build the AGI first, they will have less time to make sure that the AGI that they will build will care deeply for humans.*" ... "*Given these kinds of concerns it will be important that AGI is somehow built as a cooperation between multiple countries. The future is going to be good for AI regardless. Would be nice if it were good for humans as well.*". Last October, he told an MIT magazine interviewer that his main professional priority had changed to figuring out "*how to stop an artificial superintelligence from going rogue*".

Anthropic's CEO **Dario Amodei**, suggested in a recent interview (from 1.46.07 to 1.49.00) that, to reduce safety risks and the ongoing arms race, the development of advanced AI may have better been developed by large open intergovernmental consortia, like those that have pull tens of billions of resources together to build share large telescopes or shared atomic particle accelerators.

Last July 11th, **Google DeepMind** took a significant step by publishing a paper in collaboration with top AI scientists, introduced via a blog post, detailing very detailed "exploration" of the feasibility of creating 4 new IGOs, one of which is the Frontier AI Collaborative, an "*international public-private partnership*" to "*develop and distribute cutting- edge AI systems, or to ensure such technologies are accessible to a broad international coalition*." While the depth and scope of this paper deserve immense praise for elevating the level of discourse, it allocates a disproportionately large role to the U.S. and the U.K., and leading AI companies. This is justified based on their expertise, the need for quick action, and the urgency of certain risks.

Researchers at NGOs**,** such as research and advocacy institutes like the ***Future of Humanity Institute,*** its spinoff ***Center for the Governance of AI***, and the ***Future of Life Institute,*** have published numerous papers exploring the feasibility, complexities, and historical precedents for ambitious global AI governance of dangerous technologies. However, none have yet produced detailed proposals for such governance and processes leading up to them.

Two weeks before DeepMind's proposal, the **Trustless Computing Association** presented version 1.0 of this paper, the Harnessing AI Risk Proposal v.1, detailing a proposal for the establishment of three new IGOs for wholly managed AI, including the *Trustless Computing Certification Body and Seevik Net Initiative* for more trustworthy and widely trusted governance-support systems. It unveiled this proposal at a formal public event on June 28th at the UN in Geneva, organized by the Community of Democracies and attended by its 40 member states.

Other labs were less inclined to participatory global governance. Mustafa Suleyman, CEO of **Inflection** AI, has been extremely vocal about the need for some form of worldwide regulation. Still, he has de-emphasized the need for inclusivity, and described in an influential article with Ian Bremmer a key role of the US government and the top AI firms in designing and running such global institutions. While the Chief Scientist of **Microsoft,** Eric Joel Horvitz, signed a statement on AI risks, its president, Brad Smith stated that AI does not pose an existential threat, though warned of grave risks to safety, and called for more regulation, without specific reference to global regulation or new global institutions. Meanwhile, **Meta** - the mother company of Facebook, WhatsApp, and Instagram - has taken a very skeptical stance so far, with its Chief AI Scientist calling the warnings of AI existential risk "*preposterously ridiculous*".

Hence, not only the Lab could attract many top AI talents based on its superior mission, but it could also attract close collaboration or full participation by some leading US AI Labs and other states.

In addition, substantial **risks of near-term authoritarian political shifts in AI superpowers**, as warned (1.5 min video clip) by Joshua Bengio, could further entice top US AI labs to "internationalize" their ventures to avoid the risk of falling largely or wholly under the control of an unreliable, undemocratic or authoritarian power in the near future.

Opportunities

Find below detailed opportunities related to the *Global Public Interest AI Lab* and the *Initiative* for various entities:

- Opportunities for **States**
- Opportunities for **Donors**
- Opportunities for **NGOs and Experts**
- Opportunities for **Funders and Investors** of the Lab
- Opportunities for **Leading AI Labs**
- Opportunities for **Regional Intergovernmental Organizations**

Harnessing AI Risk Summit & Pre-Summit Virtual Conference

(this is the same text as the Summit's web page as of April 11th, 2024)

At a Glance

When & Where: The *1st Harnessing AI Risk Summit* will be held in TBD in November 2024 in Geneva, preceded by a *Pre-Summit Virtual Conference* on June 12th, 2024.

Who: A mix of globally-diverse states and IGOs. A mix of globally-diverse or neutral experts, former public officials, diplomats and NGOs in AI safety, international governance. Leading AI firms.

What: The Summit series is a key milestone of the Harnessing AI Risk Initiative, which is aggregating a few and then a critical mass of globally-diverse states to design and jump-start timely, expert, multilateral and participatory *constituent process* for the creation of an *open* global treaty-organization to jointly build and share the most capable safe AIs, and reliably ban unsafe ones.

A Unique Treaty-Making Model: The Summit series and Initiative will largely replicate, globally and for AI only, history's most successful and democratic intergovernmental treaty-making process - the one started by 2 US states convening the Annapolis Convention and ended with the US Constitution when 9 out of 13 states ratified it.

Aims - Pre-Summit:

- Consolidate and expand a *Coalition for the Harnessing AI Risk Initiative,* made up of geographically-balanced or neutral NGOs, experts, personalities and former public officials - to expand the momentum and credibility of the Initiative vis-a-vis states and regional IGOs.
- Agree on an *Open Call for the Harnessing AI Risk Initiative (v.4),* and other documents of the Initiative.
- Produce and disseminate such calls, testimonials, articles, publications and videos to promote, explain and advocate for the Initiative.

Aims - Summit:

- Achieve preliminary agreement among a critical mass of globally-diverse states on the *Scope and Rules for the Election of an Open Transnational Constituent Assembly for AI and Digital Communications* that are sufficiently participatory, resilient, inclusive and expert to expectedly lead to an intergovernmental organization that will reliably and sustainably foster the safety, wellbeing and empowerment of all, for many generations to come.
- Achieve preliminary agreement among states, investors, funders and technical partners on their participation in a democratic, partly-decentralized public-private *Global Public Benefit AI Lab* and ecosystem.

Participants - Pre-Summit

- Globally-diverse or neutral NGOs, experts and former officials and diplomats. *(We expect that many of the 39 confirmed participants to the Summit in its original date of June 12th, will participate in the Pre-Summit.)*

Participants - Summit

- State representatives from UN missions, foreign ministry of security agencies. (*We have been engaging with 7 states' missions in Geneva*)
- Leading AI labs (*We have received initial interest by 4 of the top 5 AI labs*)
- Globally-diverse or neutral NGOs, experts and former officials and diplomats. *(We expect that many of the 39 confirmed participants to the Summit in its original date of June 12th, will participate in the Pre-Summit.)*

Agenda - Summit:

Day 1 will mix 40-minute panels and 5-10 minute “lighting talks” by top experts and NGOs. Day 2 will host a wide mix of deliberative working sessions, one-way and two-way educational sessions, multilateral and bilateral meetings. See Detailed agenda below.

Agenda - Pre-Summit:

15.30 - Online Panel:
AI Risks and opportunities: the prevailing science.

16.00 - Online Panel:
Treaty-making for technological risks: nuclear, bioweapons, encryption, climate

16.30 - Online Panel:
Treaty-making for AI: the *open intergovernmental constituent assembly* model

17.00 - Online Panel:
Mitigating the risks of competing AI coalitions, AIs and AI governance initiatives.

17.30 - Online Panel:
Foreseeing and navigating complex socio-technical future AI scenarios

18.00 - Online Panel:
Open Call for the Harnessing AI Risk Initiative (v.4)

Our Greatest Risk and Opportunity

The alarm has sounded for the immense risks posed by AI, along with its great opportunities.

Since hundreds of AI **scientists**, including two of the top three, stated last March that "*Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war*", awareness of AI safety risk has been skyrocketing.

Twenty eight **states**, accruing to 80% of the world population, acknowledged in the Bletchley Declaration such safety risks, including "loss of control". Over 55% of **citizens** surveyed in 12 developed countries were "fairly" or "very" worried about "loss of control over AI". At an invitation-only CEO Summit at Yale last June, 42% of **CEOs** surveyed said they believed AI has the potential to "destroy humanity within the next five to 10 years."

The risks of AI leading to **extreme unaccountable concentration of power and wealth** - including via misinformation, surveillance, manipulation, oligopolies and biases - is just as important and urgent. Its awareness appears just as widespread among states and citizens.

Frontier AI capabilities are expected to keep **expanding five to ten fold annually**. And that's based on growth in investments and computing power alone, without accounting for AI's increasing **ability to self-improve** and multiply the productivity of its developers. A break-neck **AI arms race** among nations and firms is unfolding.

Meanwhile, seven years after the Cambridge Analytica scandal and ten after the Snowden revelations, **social media and sensitive communications** are ever more vulnerable to abuse

and control by unaccountable entities, stifling fair and effective dialogue, within and among nations, at a time when it is most needed.

Investments in AI and AI infrastructure are exploding. If successful, OpenAI's public-private **$7 trillion AI plan** to aggregate states, funders, chip makers and power providers will either (a) create an entrenched dominant global oligopoly under US control, or else (b) possibly become the seed of a safe and democratic global governance of AI that Altman has been consistently calling for - as we argue in this article in The Yuan.

If we manage to avert catastrophic risks for safety and concentration of power, by creating proper global AI governance institutions, the benefits of *human-controllable* and *humanity-controlled AI* will be astounding in terms of abundance, peace, safety and wellbeing**.**

The potential "AI pie", if we avoid the immense risks, is so enormous that rich states and people can get richer while the poor can be much better off. But success inevitably requires a fair distribution of the power in shaping our collective future in this Digital and AI Age.

As in 1946, when the US and Russia, with their Baruch and Gromyko Plans, proposed a new independent UN agency to manage all nuclear weapons stockpiles and weapons and energy research - but failed to agree - we now have a **second chance** with AI. We can **harness the risk of AI** to turn it into an unimagined blessing for humanity, and set a governance model for other dangerous technologies and global challenges.

The need for a better Treaty-Making for AI.

The agenda of states' tech diplomats is jammed with summits this year for the global governance of AI, as part of initiatives by states or IGOs. These include the 2nd and 3rd *AI Safety Summit* in Paris and Seoul, the UN Summit of the Future with its *Global Digital Compact*, *Council of Europe* treaty on AI, and AI governance initiatives by the G7 and G20.

Other key multilateral meetings will likely be held behind-the-scenes around the Guidelines for Secure AI System Development led by the US and UK national security agencies, and OpenAI's proposed "7$ trillion AI public-private consortium".

Yet, all these severely lack in representativity, inclusion, participation and timeliness.

Leading digital and AI **superpowers** appear locked in a reckless arms race - economic, military and geopolitical - over AI and AI chips, seemingly intent on hegemonizing or at best eventually splitting control.

Meanwhile, most **other nations** individually lack the political strength and strategic autonomy to table a more democratic constituent process to safeguard their economy, sovereignty, and safety in such all-important domains.

Existing **Intergovernmental organizations** - like the G7, G20, the EU, the UN, Council of Europe, OECD, GPAI, G77 - are structurally unable to lead a *democratic* global constituent process for AI governance. That's due to their lack of a mandate, lack of representativity, their closed membership and/or their statutory over-reliance on unanimity decision-making. Hence, their initiatives severely lack in multilateralism, detail, timeliness, breadth, transparency and global inclusivity, and most controlled by a handful of states.

The ***prevailing constituent methods*** *of treaty-making* being utilized are bound to result in **severely weak, fragile and undemocratic treaties** - as they did largely in past decades - due to their reliance on loose, undefined, unstructured processes, over-reliant on unanimity, that have enabled a handful or a single state to greatly and unduly influence, distort, water down or stop the process.

A Better Method for AI Treaty-Making

Hence, there is an historical opportunity for a small number of states and NGO to lead the way by utilizing - globally and for AI only - history's most successful and democratic *intergovernmental treaty-making process*, the *intergovernmental constituent assembly method* that led from the initiative of two US states to the ratification of the US Constitution by all 13 in 1787 (as argued in this blog post)

The Summit aims to be a first key step in enabling a critical mass of globally-diverse states to design and jump-start an **open and democratic global constituent process for AI**, as sketched in the *Harnessing AI Risk Initiative*, starting from for a single small states, as Trinidad and Tobago did in the 90s with the *Coalition for the International Criminal Court*.

Summit Speaking Participants

Individuals

- **Confirmed**:
- **To-be-confirmed**: (The following were confirmed for June 12th. They will be requested for confirmation for the new November 2024 date once it will be set!)
    - Rufo Guerreschi, President of the Trustless Computing Association (TCA).
    - Ansgar Koen, Global AI Ethics and Regulatory Leader at Ernst & Young. TCA Advisor.
    - Robert Trager. Director, Oxford Martin AI Governance Initiative and International Governance Lead at the Centre for the Governance of AI.
    - Kenneth Cukier. Deputy Executive Editor of The Economist, and host of its weekly tech podcast.
    - Flynn Devine, researcher on participatory AI governance methods, including research with the Collective Intelligence Project and on 'The Recursive Public'. Co-Initiator of the Global Assembly for COP26.
    - Brando Benifei, Member of European Parliament and Co-Rapporteur of the European Parliament for the EU AI ACT.
    - Mohamed Farahat, member of UN High-Level Advisory Board on Artificial Intelligence. TCA advisor.
    - Kay Firth-Butterfield, CEO of Good Tech Advisory. Former Head of AI and Member of the Exec Comm at World Economic Forum.
    - Gordon Laforge. Senior Policy Analyst at New America Foundation. TCA Advisor.
    - Marco Landi, President, EuropIA Institut. Former Group President and COO of APPLE Computers in Cupertino. TCA steering advisor.
    - Robert Whitfield, Chair of the Transnational Working Group on AI at the World Federalist Movement. Chair of One World Trust.
    - Paul Nemitz. Principal Advisor at the European Commission. Senior Privacy and AI policy expert. TCA advisor.
    - Axel Voss. Member of European Parliament and member of the Committee on Civil Liberties, Justice and Home Affairs (LIBE), and the Committee on Artificial Intelligence in a Digital Age (AIDA).
    - Akash Wasil, AI Policy Researcher at Control AI.Former senior researcher at Center on Long-Term Risk and Center for AI Safety.
    - Muhammadou M.O. Kah. Professor and Ambassador Extraordinary & Plenipotentiary of The Gambia to Switzerland & Permanent Representative to UN Organisations at Geneva, WTO & Other International Organisations in Switzerland. TCA Advisor.

- Jan Camenisch, Chief Technology Officer of Dfinity, a blockchain-based internet computer. Phd researched with 130 papers and 140 filed patents.
- Aicha Jeridi, Vice President of the North African School and Forum of Internet Governance. Member of the African Union Multi-Stakeholder Advisory Group on Internet Governance.
- Beatrice Erkers. Chief Operating Officer, Foresight Institute.
- Allison Duettmann. Chief Executive Officer, Foresight Institute.
- Lisa Thiergart. Research Manager at Machine Intelligence Research Institute (MIRI). AI Alignment Researcher.
- David Wood, President of the London Futurists association.
- Chase Cunningham. Vice President of Security Market Research at G2. Former Chief Cryptologic Technician at the US National Security Agency. Pioneer of Zero Trust. TCA advisor.
- Darren McKee. Senior Advisor at Artificial Intelligence Governance & Safety Canada (AIGS). Author of "*Uncontrollable: The Threat of Artificial Superintelligence and the Race to Save the World*"
- Sebastian Hallensleben, Head of AI at VDE, Co-Chair of the OECD Expert Group on AI (AIGO), Chair, Joint Technical Committee 21 "Artificial Intelligence" at CEN and CENELEC.
- John Havens. Exec. Dir. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.
- Philipp Amann. Group CISO at Austrian Post. Former Head of Strategy EUROPOL Cybercrime Centre.
- Ayisha Piotti. Director of AI Policy at ETH Zurich Center for Law and Economics.
- Siddhant Chatterjee, Public Policy and Governance Strategist at Holistic AI.
- Jan Philipp Albrecht, President, Heinrich Böll Foundation. Former Greens MEP. Former Minister of Digitization of the German state of Schleswig-Holstein. TCA steering advisor.
- Alexander Kriebitz, Research Associate at the Institute for Ethics in Artificial Intelligence
- David Evan Harris, Chancellor's Public Scholar at UC Berkeley. Senior researcher at Centre for International Governance Innovation (CIGI), Brennan Center for Justice, International Computer Science Institute.
- Richard Falk, professor emeritus of international law at Princeton University. Renowned global democratization expert. Chairman of the Trustees of the Euro-Mediterranean Human Rights Monitor.

- Peter Park, MIT AI Existential Safety Postdoctoral Fellow and Director of StakeOut.AI
- Pavel Laskov, Head of the Hilti Chair of Data and Application Security University of Liechtenstein
- Albert Efimov, Chair of Engineering Cybernetics, Russian National University of Science and Technology. VP of Innovation and Research at Sberbank.
- Joe Buccino, AI policy and geopolitics expert. US Defense Ret. Colonel. TCA Advisor.
- Tjerk Timan, trustworthy and fair AI Researcher. TCA Advisor.
- Roberto Savio, communications Expert. Founder and Director of Inter Press Service. TCA advisor.

Organizations

(The following were confirmed for June 12th as of April 5th. They are to be confirmed for the new December 2024 date, as the new date will be set!)

NGOs

- Confirmed:
- To-be-confirmed: (The following were confirmed for June 12th. They will be requested for confirmation for the new November 2024 date once it will be set!)
  - Trustless Computing Association
  - Machine Intelligence Research Institute (MIRI)
  - World Federalist Movement
  - Oxford Martin AI Governance Initiative
  - ETH Zurich Center for Law and Economics
  - Heinrich Böll Foundation
  - Foresight Institute
  - Artificial Intelligence Governance & Safety Canada
  - Institute for Ethics in Artificial Intelligence
  - North African School and Forum of Internet Governance
  - Europia Institut
  - Dfinity
  - StakeOut.AI

States:

- Confirmed: none
- Engaged: Last March, we held meetings in Geneva with the missions of four states to the UN, including two heads of mission (and ambassadors) and three mission's AI and digital domain experts.

AI Labs

- Confirmed: none
- Engaged: We have been in extended talks with middle executives of three of the top five US AI Labs, since November. We started to reach out to non-US leading AI Labs and NGOs in March.

Costs

- **(Apr - July 2024) Virtual Conference**:
    - $7,000 for external event service: virtual conference organizing, live streaming, video post-production and social media for youtube and social media dissemination.
    - $18,000 for staff costs: TCA President and one to-be-hired part-time junior Communications Director:
        - Preparatory content production: youtube videos on the initiative by advisors, participants and senior staff.
        - Engagement with speakers and conference content preparation
        - Engagement with states in Geneva, and various capitals.
- **(Aug- Dec 2024) Summit**:
    - $12,000 for venue and catering for two days of the Summit in Geneva
    - $13,000 for external event service: for on-site professional video production, post-production and social media for youtube and social media dissemination.
    - $25,000 for staff costs, via TCA President, one part-time junior Communications Director, and on-site event staff (event manager, assistant, and a moderator to complement 2 other volunteer ones).
- An extra $25-50,000 would enable us to cover our ultra-slim operating costs till the end of the year, to ensure focus on operations rather than fundraising. We have detailed plans on how we can use up to $5,000,000 in our Opportunity for Donors page.

Documentation

- Refer to the "Key Documents" section at the bottom of the Initiative's page, including a *55-page Executive Summary PDF* of the Initiative and the Summit.
- Refer to the Summit's webpage.

Opportunity For States

The Problem

On their own, **states and intergovernmental organizations stand powerless in the face of AI**, unable to avoid its immense risks for human safety and for concentration of power and wealth, and unable to realize its astounding opportunities.

This is the case even in larger states like Brazil, India and Germany, or large regional intergovernmental organizations like Europe.

On their own, **even superpowers stand unable to go it alone** as mitigating the proliferation and safety risks of AI are expected to be much harder than nuclear.

The Opportunity

Hence, we invite a critical mass of globally-diverse states and regional intergovernmental organizations to join in Summits, multilateral meetings and an Initiative - in Geneva - to foster a **truly multilateral and participatory process for the creation of a new open global intergovernmental organization to jointly build and share the most capable safe AIs, and reliably ban unsafe ones.**

More specifically, we invite states and regional IGOs to join the 1st Harnessing AI Risk Summit this **November 2024 in Geneva**, a *Pre-Summit Virtual Conference* on June 12th, and/or to closed-door bilateral and multilateral meetings in Geneva or via videoconference, to explore the possibility of co-leading with other states and IGOs the Harnessing AI Risk Initiative, by joining *early State participant.*

The Initiative aims to aggregate a critical mass of globally-diverse states and firms to carefully design and jump-start an *Open Transnational Constituent Assembly for AI and Digital Communications.*

Such Assembly will be mandated to draft - in a participatory, expert and timely manner - a binding treaty for a new intergovernmental organization to **reliably ban unsafe AIs, and jointly create, regulate and benefit from the most capable safe AIs.**

The mandate of such *assembly* will include the creation of an *International AI Safety Agency, an IT Security Agency,* and a $15+ billion public-private partly-decentralised *Global Public Interest AI Lab.*

The Initiative seeks to fill the wide gaps in **global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs and firms - including by the US, China, the EU, the UN and OpenAI's public-private "trillion AI plan" - and become the platform for their convergence.

A Better Treaty-Making Method

Prevailing treaty-making methods used to build new intergovernmental organizations are very ineffective and undemocratic, as demonstrated by those on climate change and nuclear weapons.

Hence, the Initiative will largely replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

Similarly, a handful of globally diverse states and IGOs can trigger a snowball to set off such a process - globally and for the all-important domain of AI - then attracting dozens of other states and eventually the superpowers.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its 1st Summit.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we are in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participant to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th, and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit, on TBD date this November.

The Global Public Benefit AI Lab

The Initiative is convening a critical mass of globally-diverse states and leading AI firms to design and jump-start an **open global constituent assembly** to create a new intergovernmental organization for AI, that includes an international *AI Safety Agency,* an *IT Security Agency,* and an open, partly-decentralized and public-private **$15+ billion *Global Public Interest AI Lab*** and its supply chain.

- The Lab will be an open, partly-decentralized, democratically-governed joint-venture of states and tech firms aimed to achieve and sustain a solid global leadership or co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures.
- The Lab will accrue capabilities and resources of participant states and firms, and distribute dividends and control to member states and directly to its citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.

- The Lab will be primarily funded via ***project finance***, buttressed by pre-licensing and pre-commercial procurement from participating states and firms.
- The Lab will seek to achieve and **sustain a solid "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms - as detailed in our recent article on the prestigious digital policy journal, *The Yuan*.

Why another Global Governance Initiative for AI?

Unfortunately, all current global and **regional intergovernmental organizations** are unfit to turn those statements of intent in an effective, inclusive and democratic global institution-building plan, because they are structurally unable to lead a democratic process to build it, due to their lack of representativity, closed membership, lack of a mandate and/or statutory over-reliance on unanimity.

The UN Secretary General António Guterres called for such an organization. He invited states to find a consensus on statements of intent via its *Global Digital Compact,* and strengthen global governance via its *Summit for the Future,* but concurrently reminded us that "***only member states can create it, not the Secretariat of the United Nations***".

Hence, an historical opportunity and responsibility for pioneering states and NGOs to lead the way for humanity as they did before, and gain very substantial economic and political benefits in the process.

Benefits for State Participants

- **Initiative**
    - **Enjoy exceptional advantages in terms of economic development,** sovereignty, safety and civil rights deriving from the joint control and ownership of the *Global Public Interest AI Lab.* This ensures reliable long-term access and control over the most advanced safe AI systems - for both their governmental and private sector needs.
    - **Radically mitigate the immense risks to human safety and for extreme unaccountable concentration of power and wealth posed by AI.** Co-lead in shaping the ethics, limits, privacy, security, safety and accountability of AI to realize its potential to bring astounding benefits to your citizens and all of humanity, in a win-win for all.

    - **Increase their leverage** vis-a-vis other global governance and infrastructure initiatives for AI by leading states and firms.
- **Summit:**
    - **Learn** about the current initiative and prospects of global AI governance of AI, about AI safety, and about the *Initiative*.
    - **Participate as a speaking participant in the Summit. "**Observer status" or remote participation is possible. Early state participants will have a guaranteed speaking slot on day 1, and will be more prominently displayed.
    - **Participate in co-designing the *Initiative***, and so therefore in the design of global governance of AI and the constituent process leading up to them. Early participants will be displayed more prominently.
- **The first 6 participant states enjoy special benefits**:
    - **Economic advantages and discounts.** Acquire a "priority option" to become one of a limited number of *Founding State Partners of the Initiative*. As such -in relation to similar states that join later - they will enjoy advantages and discounts in respect to all fees, quotas, contributions of the first three years of Initiative and Lab, and their sovereign fund participation as *Pre-seed Funders* of the Lab:
        - 70% for the 1st to the 2nd state
        - 40% for the 3rd to the 4th state
        - 20% for the 5th to the 6th state
    - **Political prominence**. More prominently included in the Summits and in the Initiative's documents and webpages. Guaranteed speaking slot on day 1 of the Summit.

Special Terms for the US and China

The above terms for participation are different for the US and China, as global and AI superpowers. They are welcome to join at any stage, yet their participation will be held in suspension until the other also joins. The first one of those two that joins will temporarily enjoy 30% higher economic and voting power advantages, which will be reduced progressively to 0% over 5 years.

More Information

- Webpages of the Initiative and its 1st Summit, next June in Geneva.
- A live-updated 55-page Executive Summary of the Harnessing AI Risk Initiative and Summit (PDF). A copy of the web pages of the *Initiative*, its *1st Summit*, the

Opportunities page - with the addition of a detailed chapter about the foreseen *Global Public Benefit AI Lab*.

- In our Opportunities web page, we offer detailed opportunities for states; for regional intergovernmental organizations; for leading AI Labs; for funders and investors in the Lab; and for donors.
- A January 2024, 33-page Harnessing AI Risk Proposal v.3 (PDF). It details the Initiative, its rationale, design of the constituent processes, and preliminary designs of the IGO and its agencies. It sets an initial framework for the Initiative's co-design with advisors, partners and Summit participants.
- A March 2024, 900-word opinion piece we authored was published 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis global AI supply chains, **OpenAI's "$7 trillion AI plan**", and the pursuit of an effective, democratic and safe global governance of AI.
- A March 2024, 1200-word blog post, we finely detailed the **methods and strategy of the constituent process** that our *Harnessing AI Risk Initiative* is pursuing to foster a democratic, safe, and beneficial global governance of AI, and how they closely replicate those that led to the US Constitutional Convention of 1787.

Participation and Registration

If interested, we'd be happy to provide any information via email or schedule a Zoom call or meeting in Geneva at your convenience. Participation in the Summit and the Pre-Summit are limited, so please confirm via email your participation as soon as possible. Individual participants' names can be communicated later via email.

Opportunities For Donors

Are you concerned about the immense **risks** of AI for yourself, your progeny and humanity? Are you excited about the astounding potential **benefits** of AI, on the other hand, if we stave off those risks? We sure are, on both counts.

A wide majority of experts, states, and world citizens agree, by now, and most agree any real solution requires **unprecedented global coordination**.

A handful of states won't be able, on their own, to control AI as they did with partial success for nuclear and bioweapons.

Nearly all states are by now well aware that, on their own, they stand powerless in the AI domain, and a new global organization to jointly develop the most capable safe AI and reliably ban unsafe ones needs to be built. But the **prevailing treaty-making models -** based on unstructured summits and unanimous declarations - are completely ineffective, as shown by those for nuclear and climate change.

For these reasons, we are leading since last March an Initiative to replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution

via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

If you agree with the above, then you may want to **donate or help us find donors** to expand our Harnessing AI Risk Initiative and its 1st Summit this November 2024 in Geneva.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as **advisors** to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGO **participants**.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we are in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participants, and more, to join as **members** of a ***Coalition for the Harnessing AI Risk Initiative*** around the joint drafting of a 400-word *Open Call for the Harnessing AI Risk Initiative* (live draft doc), *Pre-Summit Virtual Conference* starting June 12th, and started attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit in November.

Effectiveness

Not only is our strategy quite unique in its potential to create huge returns in terms of positive change, but we have demonstrated exceptional dedication, extreme frugality and high effectiveness. So your donation will be put to very good use.

As you can read in our About Us page, the milestones and momentum that we accomplished so far are the fruits of volunteer work and great personal sacrifice by Rufo Guerreschi and from 32 advisors, and 39 participants and 13 participant organizations. We received minimal non-salary funding from our spin-off TRUSTLESS.AI, till its closing last September.

Imagine what we can achieve together if properly funded!

Benefits for Donors

- Contribute to an **extremely effective way to promote the global public good**, safety and wellbeing or current and future generations.
- Contribute effectively to tackle head-on what is arguably **the greatest challenge that has ever faced humanity**: the lack of an empowered, expert, federal and democratic global governance of AI and digital communications. A solution of such challenge would:
    - Enable humanity to stave off the immense risks posed by AI and digital communications, and realize their astounding potential.
    - Protect the safety and wellbeing of your kids, and their kids, for generations to come.
    - Affirm your legacy for decades to come, being prominently recognized and publicized as an Early Patron of the Harnessing AI Risk Initiative.
- **Join a community and movement** of bright and good people.
    - Receive invitations to social events and dinners during our Harnessing AI Risk Summit and preparatory meetings in Geneva and elsewhere.
    - Join and host patron's meetings in your city.
    - Participate as a speaking participant of the Summit.
- Acquire an option to participate as an interested *pre-seed funder or investor* in the $15+ billion *Global Public Interest AI Lab*.

Operational Costs for 2024

- **Apr - July 2024 (Pre-Summit Virtual Conference)**:
    - $7,000 for external event service: virtual conference organizing, live streaming, video post-production and social media for youtube and social media dissemination.
    - $18,000 for staff costs: TCA President and one to-be-hired part-time junior Communications Director:
        - Preparatory content production: youtube videos on the initiative by advisors, participants and senior staff.
        - Engagement with speakers and conference content preparation
        - Engagement with states in Geneva, and various capitals.
- **(Aug- Dec 2024) Summit**:
    - $12,000 for venue and catering for two days of the Summit in Geneva

    - $13,000 for external event service: for on-site professional video production, post-production and social media for youtube and social media dissemination.
    - $25,000 for staff costs, via TCA President, one part-time junior Communications Director, and on-site event staff (event manager, assistant, and a moderator to complement 2 other volunteer ones).
- An extra $25-50,000 would enable us to cover our ultra-slim operating costs till the end of the year, to ensure focus on operations rather than fundraising. We have detailed plans on how we can use up to $5,000,000 in our 14-page Master Grant Proposal.

Use of Donations

- **$25-75k will enable us to power-charge our Pre-Summit and hold our 1st Summit**
- **$1.5-5 million will enable us to sustain the Initiative's operations for 3 years**, including:
    - Salary of two senior full-time staff in Geneva and president Guerreschi
    - Office space, events budget, communications, two junior staff
    - Funds beyond $1.5m will be used to outsource to partner NGOs for engagement with states in world capitals and in the Global South, and joint scientific work in AI governance and AI safety governance.

If you are Interested

Documents available on qualified request:

- A 2-page **Intro for Prospective Donors**
- A 14-page **$ 1.5-5 million Master Grant Proposal** for
- A 9-page **Case for Family Offices** donate to the *Initiative* or to invest in the *Lab.*
- A 4-page $40k *Grant Proposal for Extension of the Scientific Underpinning* (of the Initiative)

Opportunities for NGOs and Experts

We offer a globally-diverse set of leading **NGOs, scientists, academics and experts** in the fields of AI safety, AI governance, and state-grade IT security, the opportunity to - as individuals or organizations - an emerging *Coalition for the Harnessing AI Risk Initiative*, to participate in shaping and advancing such Harnessing AI Risk Initiative under the guidance of the Trustless Computing Association.

The *Initiative* is an ongoing effort aimed at aggregating a critical mass of globally diverse states to jump-start and design an **open, expert and participatory constituent process** for the creation of a new **global intergovernmental treaty-organization for AI and digital communications** that is suitable to reliably manage their immense risks in terms of human safety and concentration of power and wealth, and realize their potential to usher us in an era of unimagined prosperity, safety and well-being.

The Initiative

The Initiative aims to aggregate a critical mass of globally-diverse states and firms to carefully design and jump-start an *Open Transnational Constituent Assembly for AI and Digital Communications.*

Such Assembly will be mandated to draft - in a participatory, expert and timely manner - a binding treaty for a new intergovernmental organization to **reliably ban unsafe AIs, and jointly create, regulate and benefit from the most capable safe AIs.**

The mandate of such *assembly* will include the creation of an *International AI Safety Agency, an IT Security Agency,* and a $15+ billion public-private partly-decentralised *Global Public Interest AI Lab.*

The Initiative seeks to fill the wide gaps in **global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs and firms - including by the US, China, the EU, the UN and OpenAI's public-private "trillion AI plan" - and become the platform for their convergence.

A better Treaty-Making Method

Prevailing treaty-making methods used to build new intergovernmental organizations are very ineffective and undemocratic, as demonstrated by those on climate change and nuclear weapons.

Hence, the Initiative will largely replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

Similarly, a handful of globally diverse states and IGOs can trigger a snowball to set off such process - globally and for the all-important domain of AI - then attracting dozens of other states and eventually the superpowers.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its 1st Summit.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we are in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participant to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th, and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit, on TBD date this November.

Benefits of Member of the Coalition

As a *member* and *partner* of the Coalition, you will:

- Contribute to finalize and underwrite a 300-word *Open Call for the Harnessing AI Risk Initiative v.3* (gdoc) and future Initiative's documents.
- Optionally participate in the 1st Harnessing AI Risk Summit, this November in Geneva, and in its *Pre-Summit Virtual Conference* this June 12th on Zoom (3-7pm Geneva time).

Become a Member of the Coalition

To apply to join as *member* of the *Coalition*, takes 5-10 minutes:

- State your intent via email at info@trustlesscomputing.org. In such email:
    - state your intent to undersign the Open Call, or literal changes to it that would lead you to sign it. (Your suggested changes will be listed here)
    - state your availability to participate to the Pre-Summit and/or Summit, and future Coalition zoom calls

More Information

- Webpages of the Initiative and its 1st Summit and Pre-Summit.
- A live-updated 55-page Executive Summary of the Harnessing AI Risk Initiative and Summit (PDF). A copy of the web pages of the *Initiative*, its *1st Summit*, the *Opportunities page* - with the addition of a detailed chapter about the foreseen *Global Public Benefit AI Lab*.

- In our Opportunities web page, we offer detailed opportunities for states; for regional intergovernmental organizations; for leading AI Labs; for funders and investors in the Lab; and for donors.
- A January 2024, 33-page Harnessing AI Risk Proposal v.3 (PDF). It details the Initiative, its rationale, design of the constituent processes, and preliminary designs of the IGO and its agencies. It sets an initial framework for the Initiative's co-design with advisors, partners and Summit participants.
- A March 2024, 900-word opinion piece we authored was published 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis global AI supply chains, **OpenAI's "$7 trillion AI plan**", and the pursuit of an effective, democratic and safe global governance of AI.
- A March 2024, 1200-word blog post, we finely detailed the **methods and strategy of the constituent process** that our *Harnessing AI Risk Initiative* is pursuing to foster a democratic, safe, and beneficial global governance of AI, and how they closely replicate those that led to the US Constitutional Convention of 1787.

Opportunities for Other Entities

- Opportunities for **States**
- Opportunities for **Donors**
- Opportunities for **Funders and Investors** of the Lab
- Opportunities for **Leading AI Labs**
- Opportunities for **Regional Intergovernmental Organizations**

Opportunities for Funders & Investors in the Global AI Lab

We offer family offices, VCs, angel investors, UHNWIs, private banks, private investment funds, sovereign funds and regional intergovernmental funds (e.g. EIB, ADB) the opportunity to participate as *early pre-seed investors* in a **public-private partly-decentralized $15 billion *Global Public Interest AI Lab***.

The Lab is a key part of the Harnessing AI Risk Initiative, which is gathering in Geneva a critical mass of globally-diverse states to design and jump-start an open global constituent assembly to create a binding treaty for **a new intergovernmental organisation to reliably ban unsafe AIs, and jointly create, regulate and exploit the most capable safe AIs**.

The Initiative seeks to fill the **wide gaps in global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs and firms - including by the US, China, the EU, the UN and OpenAI's public-private "trillion AI plan" - and become the platform for their convergence.

The Global Public Interest AI Lab

- The Lab will be an open, partly-decentralized, democratically-governed joint-venture of states and suitable tech firms aimed to achieve and sustain a solid global leadership or

co-leadership in ***human-controllable* AI capability**, technical alignment research and AI safety measures.

- The Lab will accrue capabilities and resources of member states and private partners, and distribute dividends and control among member states and directly to their citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.
- The Lab will be **primarily funded via *project finance***, buttressed by pre-licensing and pre-commercial procurement from participating states and client firms.
- The Lab will seek to achieve and **sustain a resilient "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms, as detailed in our recent article on *The Yuan*.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its 1st Summit.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we are in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participant to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th, and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit, on TBD date this November.

Opportunities

- See Opportunity for investors in the Lab for more details about this opportunity to invest **$50,000 to $500,000**, and links further documentation. The funds will enable us to hire two already-identified staff with exceptional skills: a senior high-level AI infrastructure architect and a top-level Geneva-based diplomatic official. The investment carries a cap of 50x on its returns, and no voting powers except for sovereign funds.
- Participate as **speaking participants** (and possibly as sponsors) of the Summit to discuss the possibility of joining via LoIs or binding agreement as *funders or investors of the Global Public Interest AI Lab.*

More information

- Webpages of the Initiative and its 1st Summit, next June in Geneva.
- In our Opportunities web page, we offer detailed opportunities for states; for regional intergovernmental organizations; for leading AI Labs; for funders and investors in the Lab; and for donors.
- A live-updated 48-page Executive Summary of the Harnessing AI Risk Initiative and Summit (PDF). A copy of the web pages of the *Initiative*, its *1st Summit*, the *Opportunities page* - with the addition of a detailed chapter about the foreseen *Global Public Benefit AI Lab*.
- In chapter 14 of our January 2024 Harnessing AI Risk Proposal v.3 (33-page pdf), read a 2-page high-level 360° description of **the Global Public Interest AI Lab** and its feasibility. In chapter 8 of the same, read how many leading AI experts and labs have called for a "global AI Lab", and a case of why the Lab would benefit both the mission and stock valuations of participant AI labs.
- A March 2024, 900-word opinion piece I authored was published last March 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis **global AI supply chains**, **OpenAI's "$7 trillion AI plan**", and the pursuit of an effective, democratic and safe global governance of AI.
- A March 2024, 1200-word blog post, we finely detailed the **methods and strategy of the constituent process** that our *Harnessing AI Risk Initiative* is pursuing to foster a democratic, safe, and beneficial global governance of AI, and **how they closely replicate those that led to the US Constitutional Convention of 1787**.

Opportunities for Leading AI Labs

The Problem

It is becoming **increasingly unrealistic** for even the most-funded leading AI labs - on their own - to compete with BigTech for AGI leadership, secure AI market niches, or contribute to steer a reckless race for AGI on a safe and beneficial course.

OpenAI has hinted at a possible way out with its proposal for a "public-private $7 trillion AI consortium", but missed to specify that it should be open to all states and firms on equal or fair terms, and it should be democratically created and controlled.

Hence, we invite your firm to join an our Harnessing AI Risk Initiative, which is aggregating a critical mass of globally-diverse states, AI labs and key supply chain firms to build democratically a partly-decentralized public-private **$15+ billion *Global Public Benefit AI Lab,*** and an international *AI Safety Agency,* to reliably ban dangerous development and use.

We offer leading **AI labs and AI supply chain firms** the opportunity to participate as *founding innovation and go-to-market partners* of such Initiative, and start to explore such a possibility during our 1st Harnessing AI Risk Summit **this November 2024 in Geneva.**

The Initiative seeks to fill the wide **gaps in global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs

and firms - including by the US, China, the EU, the UN and **OpenAI's public-private "$7 trillion AI plan"** - and become the platform for their convergence.

The Global Public Benefit AI Lab

The Initiative is convening a critical mass of globally-diverse states and leading AI firms to design and jump-start an **open global constituent assembly** to create a new intergovernmental organization for AI, that includes an international *AI Safety Agency,* an *IT Security Agency,* and an open, partly-decentralized and public-private **$15+ billion *Global Public Interest AI Lab*** and its supply chain.

- The Lab will be an open, partly-decentralized, democratically-governed joint-venture of states and tech firms aimed to achieve and sustain a solid global leadership or co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures.
- The Lab will accrue capabilities and resources of participant states and firms, and distribute dividends and control to member states and directly to its citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.
- The Lab will be primarily funded via ***project finance***, buttressed by pre-licensing and pre-commercial procurement from participating states and firms.
- The Lab will seek to achieve and **sustain a solid "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms - as detailed in our recent article on the prestigious digital policy journal, *The Yuan*.

Better Treaty-making and Global Consortium Building

Prevailing treaty-making methods used to build new intergovernmental organizations are very ineffective and undemocratic, as demonstrated by those on climate change and nuclear weapons.

Hence, the Initiative will largely replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

Similarly, a handful of globally diverse states and IGOs can trigger a snowball to set off such process - globally and for the all-important domain of AI - then attracting dozens of other states and eventually the superpowers.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its 1st Summit.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we have been in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participant to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th, and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit, on TBD date this November.

Benefits for Early Participant AI Labs

If the Initiative succeeds in attracting a critical mass of states and funding, *participant AI labs* would **advance both their mission to benefit humanity, their stock valuations**, and retain their agency to innovate at the root and application level, within safety bounds:

- As *innovation partners* and IP providers, they would be compensated via revenue share, secured via long-term pre-licensing and pre-commercial procurement agreements from participating states and firms.
- As *go-to-market partners*, they would gain permanent access to the core AI/AGI capabilities, infrastructure, services and IP developed by the Lab.

    - These will near-certainly far outcompete all others in capabilities and safety, and be unique in actual and perceived trustworthiness of their safety and accountability.
    - They would maintain the freedom to innovate at both the base and application layers, and retain their ability to offer their services to states, firms and consumers, within some limits.
- Participant AI labs partnership terms will be designed so as to maximize the chances of a steady increase in their market valuation, in order to attract the participation of AI labs - such as Big Tech firms - that are governed by legal conventional US for-profit vehicles that legally mandate their CEOs to maximize shareholder value.

This setup will enable such labs to continue to innovate in capabilities and safety at the base and application layers but outside a “Wild West '' race to the bottom among states and labs, advancing both mission and market valuation.

Benefits of Early AI Lab Participants

The first **six AI labs** that will join as participants will **enjoy substantial economic advantages** in relation to the *Initiative* and the *Global Public Interest AI Lab,* relative to states that join later. More specifically, in respect to all revenue share, IP compensations, decision making, fees, co-investments that will be required of AI labs of similar kind in the future by the Initiative and the Lab:

- The 1st to the 2nd *lab participant* will receive a 45% premium
- The 3rd to the 4th *lab participant* will receive a 30% premium
- The 5th to the 6th *lab participant* will receive a 15% premium

More Information

- Webpages of the Initiative and its 1st Summit, next June in Geneva.
- A live-updated 55-page Executive Summary of the Harnessing AI Risk Initiative and Summit (PDF). A copy of the web pages of the *Initiative*, its *1st Summit*, the *Opportunities page* - with the addition of a detailed chapter about the foreseen *Global Public Benefit AI Lab*.
- In our Opportunities web page, we offer detailed opportunities for states; for regional intergovernmental organizations; for leading AI Labs; for funders and investors in the Lab; and for donors.
- A January 2024, 33-page Harnessing AI Risk Proposal v.3 (PDF). It details the Initiative, its rationale, design of the constituent processes, and preliminary designs of the IGO

and its agencies. It sets an initial framework for the Initiative's co-design with advisors, partners and Summit participants.

- A March 2024, 900-word opinion piece we authored was published 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis global AI supply chains, **OpenAI's "$7 trillion AI plan**", and the pursuit of an effective, democratic and safe global governance of AI.
- A March 2024, 1200-word blog post, we finely detailed the **methods and strategy of the constituent process** that our *Harnessing AI Risk Initiative* is pursuing to foster a democratic, safe, and beneficial global governance of AI, and how they closely replicate those that led to the US Constitutional Convention of 1787.

Opportunities For Regional Intergovernmental Organizations

The Problem

On their own, **regional intergovernmental organizations stand powerless in the face of AI**, unable to avoid its immense risks for human safety and for further concentration of power and wealth, and unable to realize its astounding opportunities. Even larger and more integrated ones like the European Union.

The limitations of their **mandate** and industrial capabilities in the **AI supply chain** make it impossible for them, on their own, to (1) achieve and sustain state-of-the-art AI **capabilities** and AI sovereignty and (2) have a proportional say in the creation of **global governance** institutions to regulate its safety, security and largely shape our future.

On their own, **even superpowers stand unable to go it alone** as mitigating the proliferation and safety risks of AI are expected to be much harder than nuclear.

The Opportunity

Hence, we invite a critical mass of globally-diverse states and regional intergovernmental organizations to join in Summits, multilateral meetings and an Initiative - in Geneva - to foster a

truly multilateral and participatory process for the creation of a new open global intergovernmental organization to jointly build and share the most capable safe AIs, and reliably ban unsafe ones.

More specifically, we invite states and regional IGOs to join the 1st Harnessing AI Risk Summit this **November 2024 in Geneva**, a *Pre-Summit Virtual Conference* on June 12th, and/or to closed-door bilateral and multilateral meetings in Geneva or via videoconference, to explore the possibility of co-leading with other states and IGOs the Harnessing AI Risk Initiative, by joining *early IGO participant.*

The Initiative seeks to fill the wide gaps in **global representation and democratic participation** left by global AI governance and infrastructure initiatives by leading states, IGOs and firms - including by the US, China, the EU, the UN and OpenAI's public-private "$7 trillion AI plan" - and become the platform for their convergence.

The Initiative aims to aggregate a critical mass of globally-diverse states and firms to carefully design and jump-start an *Open Transnational Constituent Assembly for AI and Digital Communications.*

Such Assembly will be mandated to draft - in a participatory, expert and timely manner - a binding treaty for a new intergovernmental organization to **reliably ban unsafe AIs, and jointly create, regulate and benefit from the most capable safe AIs.**

The mandate of such *assembly* will include the creation of an *International AI Safety Agency, an IT Security Agency,* and a **$15+ billion** public-private partly-decentralised *Global Public Interest AI Lab.*

A better Treaty-Making Method

Prevailing treaty-making methods used to build new intergovernmental organizations are very ineffective and undemocratic, as demonstrated by those on climate change and nuclear weapons.

Hence, the Initiative will largely replicate on a global basis and only for AI what is arguably **history's most successful and democratic intergovernmental treaty-making model**. That's the one that started with two US states convening of the Annapolis Convention in 1786, then to the approval of a federal constitution via simple majority in the US Constitutional Convention in 1787, and then its ratification by 9 states and then all 13 in 1789.

Similarly, a handful of globally diverse states and IGOs can trigger a snowball to set off such a process - globally and for the all-important domain of AI - then attracting dozens of other states and eventually the superpowers.

Momentum and Roadmap

So far, we have onboarded 32 world-class experts as advisors to the Association and Initiative, and over 39 world-class experts and policymakers and 13 NGOs, as *participants* in its 1st Summit.

In March, we held meetings with the **missions to the UN in Geneva of 4 states**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more. Together those states, from Africa and South America, have a population of 120 million, a GDP of $1.4 trillion, and sovereign funds of $130 billion.

In April, we received written interest from the Ambassador to the UN in Geneva of **one of the 3 largest regional intergovernmental organizations**, aggregating dozens of states. Since December, we are in extended talks with **3 of the 5 five AI Labs**, for their interest to participate in the *Global Public Interest AI Lab*.

We recently started engaging advisors and participant to build a *Coalition* (for the Harnessing AI Risk Initiative) around the joint drafting of a 300-word *Open Call (for the Harnessing AI Risk Initiative) v.3* (live draft doc), *Pre-Summit Virtual Conferences* starting June 12th, and attracting donors to power-charge our initiative.

We'll be hosting bilateral and multilateral meetings with states, IGOs and AI Labs in Geneva during the UN ITU WSIS (June 10-13th) and the UN AI for Good (May 25-29th), in advance of our 1st Harnessing AI Risk Summit, on TBD date this November.

The Global Public Benefit AI Lab

The Initiative is convening a critical mass of globally-diverse states and leading AI firms to design and jump-start an **open global constituent assembly** to create a new intergovernmental organization for AI, that includes an international *AI Safety Agency,* an *IT Security Agency,* and an open, partly-decentralized and public-private **$15+ billion *Global Public Interest AI Lab*** and its supply chain.

- The Lab will be an open, partly-decentralized, democratically-governed joint-venture of states and tech firms aimed to achieve and sustain a solid global leadership or

co-leadership in *human-controllable* AI capability, technical alignment research and AI safety measures.

- The Lab will accrue capabilities and resources of participant states and firms, and distribute dividends and control to member states and directly to its citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.
- The Lab will be primarily funded via ***project finance***, buttressed by pre-licensing and pre-commercial procurement from participating states and firms.
- The Lab will seek to achieve and **sustain a solid "mutual dependency" in its wider supply chain** vis-a-vis superpowers and future public-private consortia, through joint investments, diplomacy, trade relations and strategic industrial assets of participant states - while remaining open to merge with them on equal terms - as detailed in our recent article on the prestigious digital policy journal, *The Yuan*.

Why another Global Governance Initiative for AI?

Unfortunately, all current global and **regional intergovernmental organizations** are unfit to turn those statements of intent in an effective, inclusive and democratic global institution-building plan, because they are structurally unable to lead a democratic process to build it, due to their lack of representativity, closed membership, lack of a mandate and/or statutory over-reliance on unanimity.

The UN Secretary General António Guterres called for such an organization. He invited states to find a consensus on statements of intent via its *Global Digital Compact,* and strengthen global governance via its *Summit for the Future,* but concurrently reminded us that "***only member states can create it, not the Secretariat of the United Nations***".

Hence, an historical opportunity and responsibility for pioneering states and NGOs to lead the way for humanity as they did before, and gain very substantial economic and political benefits in the process.

Benefits for Regional Intergovernmental Organizations and their Member States

- **Initiative**
    - **Enjoy exceptional advantages in terms of economic development,** sovereignty, safety and civil rights deriving from the joint control and ownership of the *Global Public Interest AI Lab.* This ensures reliable long-term access and

control over the most advanced safe AI systems - for both their governmental and private sector needs.
    - **Radically mitigate the immense risks to human safety and for extreme unaccountable concentration of power and wealth posed by AI.** Co-lead in shaping the ethics, limits, privacy, security, safety and accountability of AI to realize its potential to bring astounding benefits your citizens and all of humanity, in a win-win for all.
    - **Increase their leverage** vis-a-vis other global governance and infrastructure initiatives for AI by leading states and firms.
- **Summit:**
    - **Learn** about the current initiative and prospects of global AI governance of AI, about AI safety, and about the *Initiative*.
    - **Participate as speaking participant in the Summit. "**Observer status" or remote participation is possible. Early state participants will have guaranteed speaking slot in day 1, and will be displayed more prominently displayed.
    - **Participate in co-designing the *Initiative***, and so therefore in the design of global governance of AI and the constituent process leading up to them. Early participants will be displayed more prominently.
- **The first 3 regional IGOs participants enjoy special benefits**:
    - **Economic advantages and discounts.** Acquire a "priority option" to become one of a limited number of *Founding Regional IGO Partners of the Initiative*. As such -in relation to similar states that join later - they will enjoy advantages and discounts in respect to all fees, quotas, contributions of the first three years of Initiative and Lab, and their sovereign fund participation as *Pre-seed Funders* of the Lab:
        - 35% for the 1st IGO, and their member states.
        - 25% for the 2nd IGO, and their member states.
        - 10% for the 3rd IGO, and their member states.
    - **Political prominence**. More prominently included in the Summits and in the Initiative's documents and webpages. Guaranteed speaking slot on day 1 of the Summit.

More Information

- Webpages of the Initiative and its 1st Summit, next June in Geneva.
- A live-updated 55-page Executive Summary of the Harnessing AI Risk Initiative and Summit (PDF). A copy of the web pages of the *Initiative*, its *1st Summit*, the

Opportunities page - with the addition of a detailed chapter about the foreseen *Global Public Benefit AI Lab*.

- In our Opportunities web page, we offer detailed opportunities for states; for regional intergovernmental organizations; for leading AI Labs; for funders and investors in the Lab; and for donors.
- A January 2024, 33-page Harnessing AI Risk Proposal v.3 (PDF). It details the Initiative, its rationale, design of the constituent processes, and preliminary designs of the IGO and its agencies. It sets an initial framework for the Initiative's co-design with advisors, partners and Summit participants.
- A March 2024, 900-word opinion piece we authored was published 13th on *The Yuan,* a prestigious Chinese digital and AI policy Journal. It frames our Initiative vis-a-vis global AI supply chains, **OpenAI's "$7 trillion AI plan**", and the pursuit of an effective, democratic and safe global governance of AI.
- A March 2024, 1200-word blog post, we finely detailed the **methods and strategy of the constituent process** that our *Harnessing AI Risk Initiative* is pursuing to foster a democratic, safe, and beneficial global governance of AI, and how they closely replicate those that led to the US Constitutional Convention of 1787.

Participation and Registration

If interested, we'd be happy to provide any information via email or schedule a Zoom call or meeting in Geneva at your convenience. Participation in the Summit and the Pre-Summit are limited, so please confirm via email your participation as soon as possible. Individual participants' names can be communicated later via email.

Unprecedented Opportunity for the Betterment of Humanity?

While being concerned and cautious is very well founded, it is vital to recognize that this and other looming catastrophic risks also present an unprecedented opportunity for the betterment of humanity, rivaling the positive transformational potential that opened after World War II.

This is essential to stimulate energetic and lucid action by good-will states, citizens and NGOs, focusing not only on the avoidance of a terrible threat but also on the potential for achievement of a much better future.

As mentioned above, one year after the creation of the United Nations, it became clear it could not prevent the spreading of nuclear and bioweapons expertise and capabilities, posing an unbearable risk of catastrophe. The UN Security Council failed in 1946 to agree on Russian and US formal proposals to mandate all members to transfer control of all their nuclear weapons arsenals and materials to a new single UN agency, which would then have a global exclusivity in research, development and management of nuclear weapons and energy.

Today, almost eighty years later, in the face of the acceleration and proliferation of a new catastrophically dangerous technology, **we have a second chance to tame and steer powerful technologies for the benefit of humanity by finally extending the democratic principle to the global level** - starting from the all-important domains of Artificial Intelligence and human communications - to establish a solid foundation for long-term human safety, dramatically reduce wealth and power disparities, and harness scientific progress to uplift all of humanity.

If successful, it is conceivable and hoped that the resulting governance, constituent process and governance-support systems will become a model for wider IGOs to manage **other dangerous technologies and global challenges**, moving closer to proper global federal democratic organizations that finally realize the promise of the United Nations.

About Us

Summary & Governance

We are a micro, hyper-neutral non-profit organization based in Geneva, dedicated to the radical increase of the safety, liberty and democratic accountability of digital communications and AI.

Since its foundation in 2015 till June 2023, our sole focus has been the Trustless Computing Certification Body and Seevik Net, aimed at progressively building consensus around the creation of new open, neutral and participatory **intergovernmental organizations** to develop and certify radically more *trustworthy* and *widely trusted* end-to-end IT systems, for **confidential and diplomatic communications**, and for **control subsystems for the most critical AIs,** social media and other society-critical infrastructure.

Our unique approach centers on a novel mix of battle-tested and time-proven *trustless* technical, socio-technical and governance systems, as specified in the Trustless Computing Paradigms, inspired by the trustworthiness paradigms of democratic constitutions, democratic electoral processes and citizen-jury systems.

On June 28th, 2023, after months of reckoning with the emergence of **immense risks and opportunities of AI** and their intersection with unregulated digital communications, we launched a Harnessing AI Risk Initiative for the creation of **three new global intergovernmental organizations** and participatory constituent processes leading up to them, to wholly govern AI and digital communications for the global public good.

Such Initiative includes the Trustless Computing Certification Body as one of the foreseen agencies of a new IGO, and was presented to a UN public event organized by the *Community of Democracies*, with its 32 member states.

According to the current Statute (pdf) and the original Founders' Meeting (pdf) minutes, the decision-making power of the association currently resides in its General Assembly and Board of Directors, both of which include three members, each with equal vote, and guided by its Steering, Scientific and Governance Advisory Boards.

Transparency and Funding

TCA has been mostly self-funded via cash and time contributions by its founders, advisors and team members.

Between 2015 and 2017, it received €35,000 in funding to organize editions of the Free and Safe in Cyberspace from EIT Digital Privacy and Security Action Line and ECSEL-JU, two EU governmental agencies promoting IT research for the public good.

From 2019 to 2023, it received about CHF 150,000 in funding in cash, time and office space from its spin-off startup TRUSTLESS.AI. Such spin-in was funded CHF 130,000 by six angel investors from Zurich, Luxembourg and Munich, and about CHF 600,000 in cash and “sweat equity” by its two co-founders Rufo Guerreschi and Alexandre Horvath.

The spin-off and TCA received office space and consulting services from three accelerator programs: Hardware.co (Berlin, 2016), Fintech Fusion (Geneva, 2019), MACH37 (McLean, 2021).

In 2021, as TCA moved its HQs to Geneva, the spin-off was turned into a "spin-in" of TCA, bound to be owned by it and its planned globally-representative participatory intergovernmental governance, in a group architecture similar in structure and intentions to that of the OpenAI, and then closed in October 2023 to simplify our structure.

Since March 2023, as our focus moved on the Harnessing AI Risk Initiative, we’ve been entirely self-funded via volunteer work. In late March 2024, we started raising $ 50,000 to 5 million in grant funding for the Harnessing AI Risk Initiative and Summit, as well as fees from state and non-state partners.

Our Story - Short Version

In 2015, we started advancing TCCB & Seevik Net via a series of research initiatives and publications, and the holding of the 1st Edition of the Free and Safe in Cyberspace conference series in Brussels.

In 2019, a startup *spin-in* called TRUSTLESS.AI which attracted private investments to build initial architecture, ecosystems, proof-of-concepts and systems compliant the TCCB (closed in September 2023).

In 2021, the *Trustless Computing Certification Body and Seevik Net Initiative* was launched and established in Geneva during the 8th Edition of the Free and Safe in Cyberspace, in its preliminary form, in the presence of top partners and personalities.

By Spring 2023, we held eleven editions of the Free and Safe in Cyberspace (in Geneva, Zurich, Brussels, New York and Berlin) with over 120 outstanding public and private participants. We aggregated world-class advisors and research partners, evolved the *Trustless Computing Paradigms*. Over 15 nation states and 3 IGOs have engaged in our events and constituent processes, or stated their interest for the Initiative.

On June 28th 2023, after months of reckoning with the emergence **immense risks and opportunities of AI** and their intersection with unregulated digital communications, we launched a Harnessing AI Risk Initiative for the creation of **three new global intergovernmental organizations** and participatory constituent processes leading up to them, to wholly govern AI and digital communications for the global public good. Such Initiative includes the Trustless Computing Certification Body as one of such organizations, and was presented to a UN public event organized by the *Community of Democracies*, with its 32 member states.

On October 18th 2023, we published a call for the convening of an *Open Transnational Constituent Assembly for AI and Digital Communications*, a call for a critical mass of globally-diverse nations to democratically and inclusively build such new organizations.

Next November 2024, we'll be aggregating a critical mass of pioneering nations, IGOs and vision-align entities to jump-start the constituent process of such organizations for the 1st Harnessing AI Risk Summit, in Geneva.