Master Grant Proposal for the Harnessing AI Risk Initiative

(version of April 11th, 2024)

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 are on both counts.

We also believe our best chance - to avoid the former and realize the latter - is to catalyze **democratic, timely, federal, open and expert** constituent processes for the creation of a new global intergovernmental organization for AI that can be expected to reliably steer AI for the global public good.

If you agree, then you may be interested in the opportunity to **donate or help us find donors** to expand our [Harnessing AI Risk Initiative](#) - and its [1st Summit](#)next June in Geneva - to convene a critical mass of states to build such organization, which will include an $15+ billion public-private *Global Public Interest AI Lab* and supply chain, open to all states to join to enjoy the most advanced safe AI.

We offer philanthropies, family offices, and wealthy individuals the opportunity to join our community and donate - from $10,000 to $5,000,000 - to help us realize a huge potential for the global public good.

Table of Contents

Two-pager Introduction for Donors

Problem and Opportunity: AI is bound to upend everything: power, society, military balance and human nature itself. In the space of one year, a majority of experts, heads of state and citizens have come to acknowledge the immense risks for human safety and concentration of power and wealth - along with its astounding potential.

Need for a treaty on AI: Given AI's inherent global nature, most agree that unprecedented global coordination is key to the challenge. Yet, given that current international institutions cannot by statute be reformed to do that, it is imperative to design and jump-start a treaty-making process that can reliably be expected to lead to an organization that can durably tackle such challenges, while mitigating the risks it may introduce in the process.

Democracy and Subsidiarity: To prevent undue concentration of power risk, clearly, such treaty-making process should be guided by the principles of democracy and subsidiarity, whereby power and governance should be at the lowest possible level: global, state, community, down to the individual. Yet, there is a huge global asymmetry in power and technical knowledge, and over 3 billion people remain illiterate and/or disconnected.

National Security Agencies: While, in the past, the national security agencies of a few states have protected humanity against the safety risks of proliferation and misuse of powerful technologies, with a mixed record, AI is expected to be much harder to control than nuclear weapons, and increasingly so. Therefore, it requires a much wider compliance by states and citizens for the extreme controls it will require.

Democratic and realist treaty-making for AI: Hence, it is imperative that we pursue, for AI, a treaty-making process solidly rooted in the democratic ideal but which also acknowledges huge asymmetries in power, wealth and literacy, while bindingly committing to reduce them in the longer term. The best way to do so, we believe, is to replicate (globally for AI and only) history's most successful and democratic treaty-making process - that which starting from two US states convening of the Annapolis Convention led to the ratification of the US constitution by all 13 - with weight voting adjusted primarily to population size and GDP. None of the current treaty-making processes even remotely follow such an approach.

Harnessing AI Risk Initiative: We've gathered a wide network of advisors and participant NGOs and experts to aggregate a few and then a critical mass of globally-diverse states to design and jump-start a Open Transnational Constituent Assembly for AI and Digital Communications. Such an Assembly will be mandated to build a new treaty-organization to

jointly build, regulate and exploit the most capable safe AI, and reliably ban unsafe ones. It will remain bindingly open to all states to join on equal terms. The initial 7 states will receive substantial economic and political advantages.

Expected Outcomes: By Summer 2025, we expect that at least 7 globally-diverse states, including two of medium or large ones with AI supply chain assets, and 3-4 leading AI labs, have agreed on a design for the Scope and Rules for Election of the Open Transnational Constituent Assembly for AI and Digital Communications, and have contributed fees in the order at least of a few USD millions in total, to jointly advance the Initiative and the planning of the foreseen Global Public Benefit AI Lab.

About Us & Team: The [Trustless Computing Association](), since 2015, has been building intergovernmental organization and technology to radically increase the trustworthiness of society-critical IT systems, via R&D, publication, spin-off, and global conference series. As more funding is acquired, work will be led full-time by TCA President Rufo Guerreschi, assisted by part-time volunteer [advisors]() and [participants NGOs and experts]() - experts in AI and IT safety and governance - as well as some of them and other identified senior experts as full-time or part-time staff.

Budget: While USD 60,000 would enable us to advance the Initiative and host online virtual events and one in-person summit, USD 1.5 million would enable us to build a proper team and turbocharge our operation over 3 years, to cover research, state engagements, events, summits, operations, and fundraising. We'll spend most of the funds beyond USD 1.5 million, and up to 5, to outsource activities to leading scientific and outreach partners around the world.

Conclusion: This Initiative puts forth a highly strategic, timely, inclusive, and expert-driven initiative to create a democratic global governance model for AI and digital communications- While highly ambitious, it addresses head-on immense risks and unlocking huge potential benefits for humanity with the chance of having a huge positive impact.

Milestones, Roadmap and Momentum

The Initiative was originally presented on *June 28th, 2023,* to the Community of Democracies and its 32 member states during a public event at the UN in Geneva, and it is the object of a [$ 1.5-5 million grant proposal]() to sustain and expand it with a coalition of globally-diverse or neutral NGOs.

We are starting off with 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.

So far, we have attracted world-class **NGOs, experts and policymakers** as participants, as you can see on the [Summit's page](#).

We are in extended talks with **3 of the 5 five AI Labs.** In March, we held meetings in Geneva with the **missions of 4 states to the UN**, including 3 heads of mission (and ambassadors) and 3 mission's AI and digital domain experts, and we are engaging 3 more.

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 work, 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. 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 in an **extremely effective way to promote the global public good**, safety and wellbeing or current and future generations.
- Contributing head-on with decent chances of success 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) 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.

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.

Context

As AI capabilities and innovations accelerate, and digital communications become ever more pervasive, humanity faces enormous risks of further concentration of power and wealth, AI misuse and accidents, and even extinction due to "loss of control" over AI.

Dominance of AI has become the key economic and military competitive factor among nations, while dominance of digital communications has become the key to globally shaping public consent and the actions of governments.

Global control over digital technologies and AI is now the key shaper of human history and central to humanity's ability to manage all threats and opportunities, including poverty, climate change, and human well-being.

A handful of states, security agencies, and firms are accumulating unaccountable power to influence AI and digital communications, stifling fair and effective digital democratic dialogue and deliberations, within and among nations. This is happening at a time when humanity needs it most to build a widely shared consensus for momentous shared decisions.

Technical Alignment of AI to humans' values, safety and wellbeing requires that all AIs and their control systems worldwide will be both *human-controllable*, to a very high and measurable extent.

Governance Alignment of AI and Digital Communications, on the other hand, requires also that both AI and digital communications will be primarily *humanity-controlled*, by individuals, nations and humans altogether, according to the *subsidiarity principle* - while safeguarding the guardian and innovative role of private initiative.

Immense Risks

The acceleration and proliferation of AI capabilities in the next few years and decades pose two main classes of risks, that of **extreme unaccountable concentration of power and wealth**, and that of **immense risks to human safety** deriving from misuse or accidents, and to loss of human control.

Frontier AIs are expected to keep expanding their capabilities 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. Meanwhile, ever more powerful AI technologies are leaked, stolen, released or reverse-engineered in reusable and modifiable formats.

A growing awareness of those risks has reached a **majority** of experts, nations, and citizens. A majority of AI scientists and experts have been sounding a loud alarm.

A statement by many AI **scientists**, including two of the top three, stated last May in an open letter, "*Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war*."

Last November, 28 **states**, amounting to 80% of the world population, recognized such safety risks, including "loss of control". A recent international **citizens**' survey in 12 countries found that over 55% are "fairly" or "very" worried about "loss of control over AI". At an invitation-only Yale CEO Summit last June, 42% of **CEO**s surveyed said they believed AI has the potential to "destroy humanity within the next five to 10 years."

"You have these trillion-dollar [AI] companies that are produced overnight, and they concentrate enormous wealth and power with a smaller and smaller number of people. …That will create a bigger and bigger distance between the haves and the have-nots, and that's another thing that causes tremendous instability in our world. And I don't know if you have an idea of how you overcome that?" stated Benjamin Netanyahu, not widely known for communist tendencies.

The safety risks are subject to high uncertainty, but some of the greatest world experts believe those could be just years away. Yoshua Bengio, recipient of the prestigious Turing Award, states that loss of control to rogue AI systems could occur in as little as a **few years** unless appropriate precautions are taken.

Anthropic's CEO Dario Amodei believes that the chance of an AI-induced civilizational catastrophe is "somewhere **between 10-25%**", while Amodei has also testified that he believes dangerous AI systems will be created soon: he claimed that systems that enable biological weapons are likely to emerge **within two to three years**. Google Deepmind's CEO Hassabis believes AGI it "could be just a **few years**, maybe a decade away."

The principle of precaution, necessarily given the immense risks of loss of human control, mandates that those warnings be taken very seriously.

Unimagined Opportunities

If we succeed in durably tackling AI's immense risks of loss of control and concentration of power, we'll have access to extremely capable *human-controllable* and *humanity-controlled* AIs to advance our own wellbeing, overwhelmingly controlled by us collectively and individually according to the *subsidiarity principle*, rather than than unaccountable entities.

Such AIs have the potential to improve human life enormously, not only in practical areas like hunger and poverty eradication, improved health, scientific innovation, cheap clean energy and increased leisure through efficiency, but also by dramatically enhancing human social, mental and emotional well-being.

If applied to our digital and social communications with controls and incentives aimed at promoting users' wellness rather than advancing the goals of political and commercial advertisers, AIs could dramatically improve communications and relationships among persons, groups and nations, promoting peace within and across nations, communities and families.

AIs could greatly improve our ability to master and transform unpleasant mental and emotional habits and patterns, as taught by spiritual traditions. Envision everyone having access to a 24/7 personal assistant, advisor, coach, doctor, wellness consultant, spiritual guide that - interacting with trusted human ones and personal data - providing tailored guidance that surpass even the best and wisest human experts, fostering significant growth in emotional intelligence, happiness and self-awareness.

These benefits could compound over time, radically raising the average happiness and wisdom of humans, advancing dramatically towards the realization of humanity's greatest potential, as pursued by our kindest, wisest and most enlightened ancestors, and possibly igniting an ever expanding happiness movement realizing humanity's greatest potential.

AI's dual potential role in materially enriching life, advancing safety and peace, and fostering better relational and inner well-being can herald an unprecedented and unimagined whole new era of human advancement.

Neglegtedness & Tractability

While there are ever more funded initiatives to tackle the technical aspects of AI alignment - like **technical** safety standards, evaluation metrics, and technical innovations that reconcile AI capabilities with its safety - the **governance** aspects of AI alignment have received much less attention, especially for initiatives that pursue ambitious truly resolutory global governance solutions and processes leading up to them.

The UN Secretary General lacks the mandate to actively call member states to create new IGOs.

Many IGOs, states, superpowers, OpenAI, Anthropic, and Google have called for the creation of participatory global organizations to manage AI risks or opportunities, yet none of them is in a condition to be a convenor or states and firms to come together to build them.

Initiatives by current global and regional intergovernmental organizations suffer from the fact that these are just fora, bound to unanimity even for press releases, and lack in global-representativity.

Hence, an opportunity for an NGO or a coalition of NGOn to act as convenors, as the World Federalist Movement did in the 1990s for the creation of the International Criminal Court.

While undoubtedly challenging and fraught with risks, building suitable powerful democratic institutions at the global level is arguably the most durable and impactful way to improve human safety and well-being in both the short and long term - as they historically did at the national level - given the current state of global semi-anarchy and the foreseen acceleration in the creation of dangerous innovations.

Efforts by NGOs to positively influence the public policy arena are hard to succeed and measure. Promoting democracy in the global governance of AI, where competence is essential, may even have adverse effects. While likely interested in joining, nearly all nations often lack the strategic autonomy to initiate such an initiative, as they could be subject to reprisals of various kinds.

This becomes a unique opportunity and responsibility for one or more NGOs to act as convenors. As stated above, OpenAI, Anthropic, UN and G77 can already be counted as supported, while many others are expected to join later as our plan will prove to be the most rational, with many win-wins even for incumbent nations and firms, as the pie will be huge.

A Second Chance

In 1946, as the world realized the immense dangers of nuclear weapon proliferation and the great promises of nuclear energy, the United States and Russia proposed the Baruch and Gromyko Plans to the United Nations for the creation of a new intergovernmental organization to centralize all nuclear weapons and nuclear energy capabilities. In the end, neither plan was ever approved by the five WW2 winners, and only veto-holding members of the UN Security Council.

Their security agencies stepped in to fill such political failure, in their ways. It was only after every one of the five WW2 winners had achieved nuclear weapons capability that an International Atomic Energy Agency (IAEA) was created in 1956, seemingly to prevent others from doing the same. Seventy years later, we are still alive, but came close to catastrophic nuclear accidents, and the risk of nuclear war is higher today than ever was.

Today, we have a second chance to do it right with AI. The stakes are even higher. Since containing proliferation of dangerous AI is expected to be harder than nuclear weaponry, pursuing a global governance plan like that attempted 1946 can succeed only if there is a wide and inclusive participation of nations and world citizens in its design.

Goals & Objectives

1. Jump-start a suitable constituent process.

The Initiative will aggregate a critical mass of globally-diverse states to join meetings in Geneva and online aimed eventually at arriving at a timely agreement on the *Scope and Rules for the Election of an Open Transnational Constituent Assembly for AI and Digital Communications*, starting with the 12th Edition of the Free and Safe in Cyberspace next June 12-13th, 2024, in Geneva. Preparatory meetings will be held in Geneva, London, Rome, and possibly Delhi and Beijing.

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. We believe, in fact, the latter is crucial to ensure that the resulting organizations will be sufficiently trustworthy and widely trusted to:

- Encourage broad adoption and compliance to the necessary worldwide bans and oversight;
- Enhance safety through global diversity and transparency in the setting of safety standards;
- Ensure a fair and safe distribution of power and wealth; and
- Mitigate the risks of destructive competition and global military instability.

2. Draft and catalyze comprehensive preliminary designs of the IGO.

The Initiative will advance - 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 such functions in three agencies of a single IGO, subject to 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 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, distribute dividends and control to member states and directly to its citizens, all the while stimulating and safeguarding private initiative for innovation and oversight.

- (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.

3. **Expert Engagement**. Engage and bring together dozens of leading globally-representative experts in AI, IT security, and global governance to join events, working groups and advisory boards to explore the complex dynamics of different models of global governance and constituent processes, from diverse viewpoints.

4. States Engagement. Engage and bring together dozens of globally-diverse nations and IGOs - and other expert, globally-representative or vision-aligned entities - in closed-door and public meetings and events, to increase knowledge sharing and consensus building.

5. Digital Dissemination. Dozens of online and in-person interviews and event sessions, with AI and governance experts, and state representatives, will be recorded and disseminated in long, short and short-clip formats, via youtube and social media.

Expected Impact

1. Ongoing Constituent Assembly: By the third year at the latest, or sooner, at least a dozen of globally-diverse nations formally and actively involved in an ongoing properly-designed *Open Transnational Constituent Assembly for AI and Digital Communications*, and a wide consensus has been built on comprehensive and detailed preliminary designs for the new intergovernmental organizations foreseen as outcome of such constituent assembly. Such constituent assembly and Rules for its election will:

- Have been and have been widely perceived as fair, resilient, neutral, expert and participatory, including via transnational citizens' assemblies and neutral academic experts.

- Ensure statutory safeguards are in place to maximize the chances that the resulting IGOs will not degenerate, unduly centralize too much power or be captured by one or a few nations, and improve over time, such as via check and balances, a solid federal structure, regular mandatory re-constituent assemblies and other measures.

- Have established an open dialogue and convergence paths with competing or complementary initiatives, especially those led by digital superpowers, positioning as a medium-term complement for possible more urgent less multilateral initiatives to tackle urgent global safety risks.

2. Neutral Expert Community. Dozens of esteemed experts, partners and former diplomats have boarded via active and public contributions or partnership in the initiative, via public statements and appearances, and participation in advisory committees. Twelve or more joint white papers, academic papers and outreach materials aimed at the general public and heads of state.

3. Perceptions. Widespread concerns that creating new highly empowered global institutions inevitably leads to a concentration of power away from world citizens have been successfully countered, by making clear that doing it right will result in the opposite.

About Us

The Trustless Computing Association is a project-driven Swiss non-profit organization based in Geneva, dedicated to the radical increase of the safety, liberty and democratic accountability of digital communications and AI.

Since our foundation in 2015 till June 2023, our sole focus has been the Trustless Computing Certification Body and Seevik Net, an Initiative aimed at progressively building consensus around the creation of new open, neutral and participatory **intergovernmental organization** 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 citizens-jury systems.

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 (closed in September 2023) which attracted private investments to build initial architecture, ecosystems, proof-of-concepts and systems compliant to the TCCB.

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 have 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 month of reckoning with the emerging **immense risks and opportunities of AI** and their intersection with unregulated digital communications, we presented the Harnessing AI Risk proposal v.1 - to a UN public event organized by the *Community of Democracies*, with its 32 member states - for the creation of **a new global intergovernmental organization** with three agencies, including the Trustless Computing Certification Body, and a participatory constituent process leading up to it, to wholly govern AI and digital communications for the global public good.

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

Next June 12-13th, 2024, we'll be aggregating a critical mass of pioneering nations, IGOs and vision-align entities to jump-start the constituent process of such organization during the 12th edition of Free and Safe in Cyberspace, in Geneva.

Team & Partners

The team will be led full-time by Rufo Guerreschi, the founder and president of TCA. A number of current and future advisors of the Trustless Computing Association, and participants to our Free and Safe in Cyberspace conferences, will be supporting part-time in content writing, expert community coordination, outreach, research and networking.

- **Staff**. We have identified two full-time senior management, keenly aligned to the mission, that will be hired as we'll funded:
  - an experienced Geneva-based full-time executive to engage in one-to-one engagement with states and their missions in Geneva; and/or
  - a senior hardware/software/network architect from a frontier AI lab to act as CTO of the initiative to guide content writing and discussions with states about AI safety standards and supply-chain design of the Global AI Lab.
- **Partners**: A large part of the funds raised in excess of USD 2 million will be utilized to subsidize relevant activities by a **globally-balanced set** of highly specialized and mission-aligned NGO partners. These are some NGOs that we have identified as fitting, and started engaging, though no formal agreement with them is yet in place:

  - [TBD Geneva] for outreach, educational activities and events organization targeted at states and their mission in Geneva.

  - [TBD Chinese] for engagements with organizations in China.

  - [TBD US]for engagement with organizations in the US.

  - [TBD Middle Eastern] for engagements with organizations in the Middle East.

  - [TBD African] for engagement with organizations in Africa and the Global South.

  - [TBD participatory process NGO] for structuring participatory process for the assembly and IGO, including involving random-sampled citizens, states and possible application of global *collective constitutional AI* methods.

  - [TBD NGOs such as the Future of Humanity Institute, Centre for the Governance of AI, Oxford Martin AI Governance Initiative and/or Future of Life Institute] for joint technical, socio-technical and governance research, analysis and publications, and engagements with relevant UK/US organizations.

Budget

We estimate the total cost of the initiative to be **USD 1.5 - 5 million** over a period of 3 years:

- Research and Development: Subsidize several part-time researchers and experts to engage in publications, forums, and community development to expand the number of those that are engaged, co-authoring or contributing to publications, joining committees, work-groups and advisory boards.

- Conferences: Logistic and promotional activities related to hosting a series of conferences, and online and offline preparatory meetings, to be hosted in Geneva, including: venue, facilitation, moderation, engagement, video, promotion and marketing.

- <u>Senior Management</u>: Sustain at least one experienced Geneva-based full-time executive, in addition to the president Rufo Guerreschi, successfully engage in 1-to-1 engagement with states and their missions in Geneva, and accordingly evolve high-level content in the form of online content, proposals, engagement agreements.

- <u>Operations</u>: Sustain a cheap flexible office space in Geneva, at a coworking space near the offices of UN agencies and missions of member states to those offices. Sustain one full-time executive assistant, a few volunteers, and part-time administrative, legal consultants, digital systems administrators, and social media marketing consultants.

- <u>Fundraising</u>: Additional funds will be pursued via donations by a balanced mix of states, IGOs, philanthropies, firms, and others. Our ultimate funding aim will be for funding to come predominantly or exclusively from a globally-representative set of states via annual fees.

Budget of the Trustless Computing Association - 2024-2027

(values are in thousands of US dollars)

	2024			2025				2026				2027
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
events, summits, meetings	30	0	30	0	30	0	30	0	30	0	30	0
research & development	10	10	10	10	10	10	10	10	10	10	10	10
full-time and part-time staff	30	50	50	50	50	50	50	50	50	50	50	50
partners and advisors	30	30	30	30	30	30	30	30	30	30	30	30
marketing, dissemination	15	15	15	15	15	15	15	15	15	15	15	15
office costs	10	10	10	10	10	10	10	10	10	10	10	10
admin & legal	8	8	8	8	8	8	8	8	8	8	8	8
unforeseen	7	7	7	7	7	7	7	7	7	7	7	7
expenses	**140**	**130**	**160**	**130**	**160**	**130**	**160**	**130**	**160**	**130**	**160**	**130**
Initiatives' partner fees				100				300				500
donations			15	100	20	20	20	150	25	25	25	200
revenue	**0**	**0**	**15**	**200**	**20**	**20**	**20**	**450**	**25**	**25**	**25**	**700**
balance to-date	1,500	1,370	1,225	1,295	1,155	1,045	905	1,225	1,090	985	850	1,420

Conclusions

This grant proposal outlines an highly ambitious but achievable plan to substantially contribute to foster and catalyze the creation of a suitable democratic global governance of AI and digital communications for humanity's safety and well-being. A strong team of experts supports the proposal and has a realistic budget. I urge you to fund this important initiative.

More Information

For more information, after reviewing the Harnessing AI Risk Initiative and the June 2024 Summit web pages, and the Summit's participant list, review in order:

- **Our website,** for more about us, our team, advisors and partner, and our offers to join for other stakeholders.
- **Harnessing AI Risk Proposal v.3**, a 33-pager text detailing the Initiative, its rationale, design of the constituent processes, and preliminary designs of the IGO and its agencies. It set an initial framework for the Initiative's co-design with advisors, partners and Summit participants. (Jan 15th, 2024. pdf preprint on ResearchGate).
- **Can OpenAI's "$7 trillion plan" become the seed of a global, federal, safe and democratic ecosystem, lab and safety agency for AI?**, a blog and linkedin post that analyzes such a plan in the context of Altman's repeated calls for a participatory global constituent assembly for AI, how it relates to our Initiative. (Feb 14th, 2024, blog post and linkedin post).
- **Case for family offices to support and invest in a Global Public Benefit AI Lab and an International AI Safety Agency,** a 8-page case, preceded by a 2-page summary (Feb 2024, pdf).

Contacts

Rufo Guerreschi, President of the Trustless Computing Association
rufo@trustlesscomputing.org — mobile +393289376075 — WA/Signal +41799137280