

# Case for family offices to support and invest in a Global Public Benefit AI Lab and an International AI Safety Agency (v.1)

(Published on February 8th, 2024)

by Rufo Guerreschi - Trustless Computing Association

**Abstract:** This text argues that family offices could increase patrimony, effectiveness of their philanthropy, and wellbeing of future generation, by supporting or investing in the participatory creation of a USD 15+ billion Global Public Benefit AI Lab and an Intl' AI Safety Agency - via the [Harnessing AI Risk Initiative](#) - to jointly create, regulate and benefit from the most capable safe AIs, and reliably ban unsafe ones. While focused on family offices, this case is largely fitting for philanthropies, private banks and sovereign and pension funds.

## Executive Summary

Wealthy individuals and their family offices have greatly benefited overall from the digital revolution so far, even more than ordinary citizens.

Their patrimony benefited from secure and profitable investments in Bg Tech and the startups they regularly acquire. Their family firms benefited from increased productivity. Their philanthropy was able to foster effective tech-driven development and digitization for the less privileged. Their children enjoyed increased abundance and opportunities.

More recently, fast-improving AI is starting to help family offices by improving their investment decisions, risk management, estate planning, operations automation, and by enhancing the digital user experience. In the next few years, it is widely expected to bring **unimagined opportunities** improvements in productivity and quality of their services. The prospects are astounding.

In a [recent video interview](#) in Davos 2024, the CEO of Inflection AI, one of the top 5 AI labs, described PI, their multi-billion dollar AI assistant, as basically a digital family office of everyone, "*Hundreds of millions of people who do not have access to kindness and support, to a good analyst, to a chief of staff, to a scheduler, to a financial advisor. You know these services are only available to the top few percent of us in the world today. In the future PI is going to be like having a high-net-worth family office for everybody*".

The value of the top ten Big Tech companies grew from about USD 5 trillion five years ago to 9 trillion today and is expected to be valued at 21 trillion in 5 years. With a current 14x ratio between revenue and stock valuation, their foreseen valuation is primarily based on the expectations they'll further entrench their market positions in AI, social media and digital communications, trumping the role, power and wealth of democratic institutions.

Yet, this further expansion of digital innovation and Big Tech, and especially the advent of advanced AI, comes with both **immense risks** for family offices in the next few years.

The shocking acceleration of AI innovation and investment, and its convergence with digital communications poses very real, pose substantial short or medium term risk of **extreme concentration of**

**power and wealth in a few individuals and states**, catastrophic risks for safety, privacy, geopolitical stability, and even unimagined catastrophe due to **loss of human control** over AI.

As power and wealth is accruing to a few dozens individuals and financial institutions that control such Big Techs, family offices are more and more pressed, as are sovereign funds, to get on the "Big Tech stocks bandwagon", to increase and preserve their patrimony, even though it often seems such bandwagon is heading towards a precipice dragging all of us with it. At times there seems to be too many hands on the wheel. At other times, no hands on the wheel at all.

While the economic and geopolitical power dynamics at play around AI may seem unstoppable, we believe there is an **opportunity for a critical mass of family offices to join forces with a coalition of globally diverse states, sovereign and pension funds, and responsible AI firms, to steer this revolution in the interest of all of humanity**, while safeguarding the wellbeing of future generations, protecting and increasing their patrimony.

A critical mass of globally-diverse states, and open to all other states to join on equal terms, could agree on and jump-start a democratic, timely and expert constituent processes to build an empowered global intergovernmental organization to jointly create, regulate and benefit from the most capable safe AIs and reliably ban unsafe ones.

Such initiative should include *Global Public Benefit AI Lab* with an endowment of at least USD 15+ billion that will set up slated to do for AI what the USD 7 billion Airbus did for civil aviation and the USD 30 billion International Thermonuclear Experimental Reactor (ITER) is doing for nuclear fusion energy.

Given the proven value and scalability of the most advanced LLMs, and the fact that it will be buttressed by pre-licensing and pre-commercial procurement contracts with participant states, IGOs and large private organizations, it will be primarily funds via *project finance*, providing relatively solid and predictable returns to state-owned and private investment entities.

Such are the aims of the [Harnessing AI Risk Initiative](#) that our [Trustless Computing Association](#) has launched in Geneva since last June, which will be jump-started with its [1st Harnessing AI Risk Summit](#) next June 12-13th, 2024, in Geneva, already adhered to by world-class participants.

Family offices could participate as patrons of such an initiative, and as early investors in such a Global AI lab, advancing dramatically all their main concerns and interests, as well as the wellbeing and safety of all of humanity, as we detail below.

## Table of Contents

*Executive Summary*..... 1  
*Table of Contents*..... 2  
*Family Offices and the Digital and AI Revolutions*.....3  
*Safety and Hope of Future Generations*.....3  
*Hyper-Concentration of Power and Wealth*..... 4  
*Privacy, Liberty and Agency*.....4  
*Global Governance of AI and Digital Communications*.....4  
*Prospects and Need of a Democratic Global AI Lab*..... 5  
*Philanthropic and Investment Opportunities*.....7

## Family Offices and the Digital and AI Revolutions

Wealthy individuals and their family offices are constantly faced with the complexities of understanding and confronting the unfolding revolution in AI and digital communications and their impact on their patrimony, personal privacy and freedom, the safety of future generations and their philanthropic activities.

They are rightly proud of their role in advancing economic development, creating jobs and abundance for all, safeguarding future generations, and using part of their wealth to help alleviate suffering and affect positive change in the short and the long term.

The shocking recent acceleration in AI capabilities has the potential to usher humanity into an era of unimaginable abundance, privacy, safety and human well-being and to empower family offices to advance their interests and values.

In this article, we will look at the main threats for family offices posed by digital technologies, understand their foreseen dynamics, explore mitigations, and argue how supporting in a wide variety of ways suitable global governance may be the most effective way to turn those immense threats into unprecedented opportunities for them and humanity at large.

## Safety and Hope of Future Generations

It has become the majority scientific consensus that AI, unless mitigated, poses substantial existential risks to the life and livelihood of the owners and clients of family offices, their future generations, and the rest of humanity, akin to and beyond the threat posed by the proliferation of nuclear weapons after World War II.

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.*"

OpenAI CEO Sam Altman has stated that the development of superhuman machine intelligence is "*the greatest threat to the continued existence of humanity,*" and Anthropic 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 considers dangerous AI systems will be created soon: he claimed that systems that enable biological weapons are likely to emerge **within two to three years**.

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.

Last November, 28 states, amounting to 80% of the world population, recognized safety risks in the [Bletchley Declaration](#), including the risk of "loss of control." At an invitation-only Yale CEO Summit last June, 42% of CEOs [surveyed](#) said they believed AI has the potential to "*destroy humanity within the next five to 10 years.*"

A recent UK [survey](#) of citizens in 12 developed countries found that over 55% of them are "fairly" or "very" worried about "loss of control over AI." The same survey found that a majority of **young people**, while more optimistic than older people, believe that *"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."*

These facts are a clear sign that new generations are burdened with **profound risks and deep fears about the future** that the current trajectory of AI will bring, which compounds those about climate change and nuclear war.

## **Hyper-Concentration of Power and Wealth**

The patrimony and agency of a vast majority of family offices and (ultra)high-net-worth and patrimony are also greatly threatened.

The current dynamics at play are leading to a rapid extreme and unaccountable hyper-concentration of power and wealth in the hands of very few superpowers, individuals, and giant investment banks that control a few big tech and new frontier AI startups.

Such concentration of power is happening in an uncoordinated way, in a breakneck arms race among a dozen firms and the superpowers that back them, which causes each of them to charge forward irresponsibly, underplaying the immense risk for safety and loss of control.

Most value in the markets is accruing in a few leading Big Tech and AI companies and a handful of giant institutional investors, slated to surpass tens of trillion dollars in valuation if the current trend continues. Those giant financial institutions and Big Tech that control and will control the most advanced AIs use their privileged access to powerful AI to increase their ability to predict and manipulate the markets to their advantage, further leaving everyone else behind.

Traditional family businesses and family office assets are expected to lose value over time as the whole economy is, with an ever-faster acceleration, eaten up by those few giants and their principal shareholders.

## **Privacy, Liberty and Agency**

Their privacy and liberty are foreseen to be reduced even more than before.

Along with influential politicians and investigative journalists, high-net-worth individuals and family offices suffer the most in the invasion of privacy by states and criminals due to their nature as valuable targets. Paradoxically, they have much less privacy than a bus driver, whose personal secrets or whereabouts are worthless to criminals or rogue states.

Given the need for governments to spy on the bad guys - and the absence of legal access mechanism for digital communications, as opposed to analog phone and paper mail - all commercially available systems and standards are systematically and surreptitiously compromised by powerful states, making them vulnerable also to innumerable criminal entities and rogue states, as collateral damage.

AI, of course, given its control by a handful of states and firms seeking to profit from our data and increase control, and how intrusive it will become in our personal lives, is slated to make this situation much worse.

## **Global Governance of AI and Digital Communications**

Avoiding the mentioned risks and realizing the opportunities for family offices and humanity requires primarily governance solutions rather than technical ones, as only proper governance can ensure that research funds are appropriately directed and its results shared and controlled when dangerous.

Several efforts are underway - funded by states, leading firms and NGOs financed by a handful of US philanthropies and family offices - to solve the "**technical half**" of **AI alignment**, find technical and socio-technical breakthroughs to ensure that the most potent future AIs will somehow remain under human control and align with the interests of humans, rather than sacrifice pursue some other unforeseen goals, even if that means severely limiting AI capabilities.

Even if they succeed, far far from certain, the "**governance half**" of the **AI alignment problem** remains, which entails two key challenges: (1) how to prevent others anywhere in the world from irresponsibly developing AI that is not "technically aligned" and (2) how to prevent such success would lead to an immense concentration of power in the entity that succeeds.

For these reasons, a consensus is widely spreading that there is a need for **participatory, federal and empowered global governance mechanisms** that can steer us away from the ongoing arms race among states and firms, reliably manage its immense and urgent risks, and ensure their control and benefits are sufficiently shared among nations and world citizens, in a win-win post-scarcity scenario.

Yet, proper global governance initiatives are missing. The current ones are severely lacking in scope and detail and are highly fragmented.

AI **superpowers** China and the US appear locked in an economic and military AI arms race, driven by narrow interests and entrenched in a rhetoric of mistrust. They do not show serious shared plans to avoid AI safety risks nor provide concrete guarantees about the equal sharing of the enormous benefits and power that will derive from AI.

**Inter-governmental organizations** like the EU, UN, G7, G20, and G77 have issued statements of principle but have proven once again to be just forums, unable even to propose what's needed due to their unanimity and veto-based decision-making.

Meanwhile, other **nations** lack the political strength and strategic autonomy, alone, to table alternative proposals in such all-important domains.

## Prospects and Need for a Global Public Benefit AI Lab

Given the *generalist* nature of the frontier AIs being developed, lack of effective antitrust laws, network dynamics and the huge investments needed to stay on top, it is very hard for non-superpower states or their firms to keep pace and avoid becoming more and more "digital colonies."

Following on the trails of [the UK](#), [Israel](#), [UAE and Saudi Arabia](#), many EU and non-EU states are seeking to build their own national advanced AI capabilities to safeguard and promote their economic development and global competitiveness, as well as sovereignty and national security. In France, four labs were funded, and the entrepreneur Xavier Niel has [raised](#) a puny €300 million for an initiative branded as "national," backed by the former president of Microsoft.

The German AI Association has been [promoting](#) a €400 million LEAM initiative. German and French firms, like Mistral and Aleph Alpha, have raised a few hundred million to create frontier AIs based on EU values and are seeking more. A [Swiss AI Initiative](#) was announced to build a frontier AI model infrastructure made of 10,000 next-generation GPUs.

None of these are remotely a match for the technological investment and market access of BgTech. Meta [declared](#) it will invest \$10 billion in specialized chips for LLM training, while Microsoft is investing similar amounts in AI infrastructure and chips. Anthropic, OpenAI and Deepmind are slated to invest billions next year in "training" the next generations of AI, and likely tens of billions in a few years if steady trends of the last 10 years continue.

Given the highly generalist nature of AI technologies and network effects, it is implausible that national AI initiatives can offer services that are remotely competitive with those of the giants, even to their own governments, firms, and citizens. In addition, while supported by their states as "sovereign" champions, they depend on chips, compute power and capital controlled by the current AI superpowers.

Last but not least, most of those initiatives are mainly basing their claim of future competitiveness on vague and partial **claims of transparency and "open-source" to attract talent** and stake claims of trustworthiness and safety that open-sourcing some components do not remotely warrant. Such a strategy is bound to fall when open sourcing LLM model weights and then other AI techs will likely be globally regulated with regimes similar to nuclear proliferation, as a large majority of AI scientists and experts foresee.

Hence, as it stands, the **only way to profit for the family office seems to be to come in as a late investor in oligopolistic US giants**, engaged in an irresponsible arms race that could lead us to extinction, and a profound gutting of democracy and liberty worldwide through concentration of power, but whose share price is slated to grow steadily.

While the possibility to invest in European digital or AI champions trying to compete faces a lack of market and investments large enough, investments in smaller AI startups face the huge risk that whatever capabilities and market they create will be eaten up by the next generation of general models of global giants.

So, family offices face the dilemma of investing in irresponsible digital giants to safeguard and increase their patrimony, even though they are slated to damage them and their progeny immensely in the medium and long term.

Early ideas are being floated by the *UN High-Level Advisory Body on Artificial Intelligence* via its December 2023 [Interim Report](#) for a critical mass of states to create some sort of global AI lab similar to the USD 30 billion International Thermonuclear Experimental Reactor (ITER) for nuclear fusion energy, or Airbus for civil aviation. Such ideas are supported by calls by top leading AI scientists, AI Labs and NGOs for multi-national or global AI labs, a "CERN for AI."

The [Harnessing AI Risk Initiative](#) by the [Trustless Computing Association](#) brings such ideas further, proposing the creation of a USD 20+ billion *Global Public Interest AI Lab*, a democratically-governed public-private initiative, controlled by an open globally-diverse critical mass of states, financed primarily via a *project finance* model, bolstered by pre-licensing and pre-commercial procurement contracts with participating states and private organizations, and funded primarily by sovereign and pensions funds, family offices, large private banks.

Such a sizable public-private infrastructure - similar to large energy and transportation infrastructure projects - could become an investment option that produces returns for family offices comparable to and higher than investing in BigTech, with less risk, and extensively promoting family offices and humanity's interest.

## **Philanthropic and Investment Opportunities**

Several traditional philanthropic entities supported by US family offices and wealthy US individuals have in recent years funded NGOs, primarily based in the US and UK, to do crucial work on these problems, mainly from the technical perspective, much less on the governance side.

A handful of them, primarily funded by a handful of former Silicon Valley tech founders, and developed around the *effective altruism* movement, have donated hundreds of millions to NGOs promoting long-term AI safety. While these have done crucial, vital and much needed work, they have supported mostly technical mitigations, focused their work and communications possibly too much on very long-term risks and opportunities rather than more short term ones and are yet to support any truly multilateral international governance initiatives.

It has transpired that their vision of global governance of AI and digital communications reflects US and US West Coast mentalities and biases - which appears to have little hope or belief in a positive future of AI via globally participatory, democratic and multilateral mechanisms - and seem to assume instead an inevitable future whereby control is bound to accrue in one or two superpowers, a few US and Chinese firms, and a tight-knit West Coast intelligentsia.

The issue of digital privacy suffers from similar deficiencies, with NGOs funded mainly by private US entities based on West Coast mentalities based on either US-centric "realist" approaches or cyber-libertarian attitudes whereby privacy and public safety are framed as an "either-or" choice. The result has been a steady erosion of privacy for all citizens, especially family offices.

Hence, there is an opportunity for family offices across the world to contribute to shape the future of AI, radically advance their philanthropic activities and increase their patrimony - as well as their privacy, liberty, agency and the well-being of the next generations - by donating to NGOs that are effectively promoting democratic global governance of AI and digital communications.

## Conclusions

We hope to have argued convincingly in this text that family offices, and loose grouping of them, could advance dramatically all their main concerns and interests - as well as the wellbeing and safety of all of humanity - by supporting large impactful initiatives aimed at a participatory, expert and timely process to create coherence global governance of AI and a global AI lab, by participating as patrons of such initiatives, and as early investors in such an AI lab.

Such are the aims of the [Harnessing AI Risk Initiative](#) that our [Trustless Computing Association](#) has launched in Geneva since last June, which will be jump-started with its [1st Harnessing AI Risk Summit](#) next June 12-13th, 2024, in Geneva, already adhered to by world-class participants.

## More Information

More information on this opportunity for family offices can be found at the web pages of the initiative and Summit. Several linked documents at the bottom of the Initiative page, including a 33-page [Harnessing AI Risk Proposal v.3](#), that details in chapter 4 and 12 the global AI lab, and our 14-page [USD 1.5 - 5 million grant proposal \(pdf\)](#). See on our website for more [about us](#), our [team, advisors and partner](#), and our [offers to join](#) the Initiative to other stakeholders.

## Contacts

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