



# Annual Report

2023



# Message from the President and Executive Director

Dear Friends of FLI,

2023 marked a watershed moment in humanity’s relationship with artificial intelligence – and in FLI’s mission to ensure this transformative technology benefits rather than endangers life. Our March open letter calling for a pause in advanced AI development garnered over 33,000 signatures and catalyzed a global conversation about AI risk that reverberated through boardrooms, legislative chambers, and international summits.

The rapid evolution of AI capabilities this year validated both our long-standing concerns and the urgent need for coordinated global action. We were heartened to see this urgency reflected in strengthened EU AI regulations, unprecedented U.S. Congressional hearings, and most notably, the historic Bletchley Park AI Safety Summit – where 28 nations acknowledged the potential for catastrophic harm from advanced AI systems.

Beyond our high-profile advocacy, FLI continued its vital work across multiple fronts: Our policy team shaped crucial AI governance frameworks in both the EU and U.S. Our Futures program launched impactful projects such as “Artificial Escalation”, highlighting AI-nuclear risks through compelling storytelling. Through our grantmaking, we supported critical research in AI safety, including our expanded Vitalik Buterin Fellowship program.

The Future of Life Award honored the creators of “WarGames” and “The Day After”, recognizing how powerful narratives can drive positive change – a lesson particularly relevant as we work to help humanity grasp both the promise and peril of advanced AI.

We’re deeply grateful to our growing community – now including over 40,000 newsletter subscribers and 10,000+ podcast listeners – who share our commitment to ensuring technological progress enhances rather than endangers humanity’s future. Your support enables us to maintain our independence and continue speaking necessary truths about the challenges we face.

Looking ahead, we remain focused on our core mission: steering transformative technologies toward benefiting life and away from extreme risks. The acceleration of AI development makes this work more crucial than ever. We invite you to review our key initiatives and impacts detailed in this report, and join us in this vital effort to safeguard humanity’s future.

With determination and hope,



**Anthony Aguirre**  
Executive Director,  
Future of Life Institute



**Max Tegmark**  
President,  
Future of Life Institute

# Mission

**The Future of Life Institute's mission is to steer transformative technology towards benefiting life and away from extreme large-scale risks.**

We believe that the way powerful technology is developed and used will be the most important factor in determining the prospects for the future of life. This is why we have made it our mission to ensure that technology continues to improve those prospects.

In pursuit of this mission, we support the development of institutions and visions necessary to manage world-driving technologies and enable positive futures. We seek to reduce large-scale harm and existential risk resulting from accidental or intentional misuse of transformative technologies.



# How We Work

Towards these aims, FLI engages in a number of initiatives and strategies, including:

## Policy

Our **Policy team** works to improve governance of transformative technologies through research and advocacy. We seek to bridge the gap between the experts who understand transformative technologies and the public institutions with the legitimacy and means to govern them.

## Outreach

Our **Outreach** efforts work to help policymakers, technologists and the general public understand the challenges and opportunities we face. We work with press, media organizations, content creators, influencers and more to help drive awareness and understanding of the risks, opportunities and solutions around transformative tech. To this end we place op-eds, comment in press, conduct interviews, secure event speaking spots, guest on podcasts, create videos, and more. We also produce “homegrown” content to publish on our own channels. For example, this year, we relaunched our namesake [podcast](#) series, which features the FLI team in conversation with prominent researchers, policy experts, philosophers, and a range of other influential thinkers.

## Futures

Our **Futures** program aims to help humanity aim towards the beneficial outcomes made possible by transformative technologies. By employing tools like storytelling, worldbuilding, scenario planning, research, and forecasting, we explore possible futures and determine actionable pathways. We then pinpoint the essential policies, decisions, and institutions needed to navigate these paths, and how best to deliver them. The program seeks to engage a diverse group of stakeholders from different professions, communities, and regions to shape our shared future together.

## Grantmaking

Our **Grantmaking** provides financial support for promising work aligned with our mission. Tackling extreme, large-scale risks takes a community. We believe this area is extremely under-resourced and we see tremendous opportunity for increased grantmaking activity going forward.



# Policy

In 2023, FLI engaged with policymakers and helped shape important legislation in both Europe and the US. We contributed to many policy documents and responses, raised our organizational profile, and positively influenced US and EU policy development.

FLI has been deeply involved in advocacy surrounding the **EU's AI Act**. This legislation is poised to be the first major AI regulation passed and will include coverage of general purpose AI models.

In October 2023, Max Tegmark, FLI's President, was invited to address the **AI Insight Forum on AI Innovation** at the US Senate. His recommendations included protecting against catastrophes that could derail innovation, developing and mandating rigorous cybersecurity standards, establishing a centralized federal agency, subjecting developers of advanced general-purpose AI systems to liability for harms caused by their systems, and increasing federal funding for research and development.

In our Lethal Autonomous Weapons work, we intensively advised and supported the Costa Rican government in creating a successful blueprint for **regional conferences on autonomous weapons**. The conference resulted in a unified Latin America position in support of a treaty. The conference triggered a wave of successive conferences in Luxembourg, Trinidad and Tobago, the Philippines, and those upcoming in Sierra Leone and Austria.



## Highlight: Attending and providing input on the first ever government-convened AI Safety Summit

In a major step towards greater international collaboration, the United Kingdom convened the first ever global government meeting centered on AI Safety. The summit brought together 28 countries, leading civil society groups, AI companies, and research experts at historic Bletchley Park. FLI was heavily involved in the lead up to the summit. In September, we released our recommendations, which included a draft declaration on AI safety and a post-Summit roadmap. Impressively, the Summit culminated in a joint communique from all 28 governments warning of the potential for catastrophic harm from AI. We were also heartened to see one of our key recommendations - reconvening every six months - was followed, with AI Safety summits scheduled in South Korea and France.

FLI's President, Max Tegmark, was invited to be one of the hundred summit attendees. Speaking of the experience, Max said, "I've been working for about 10 years, hoping that one day there would be an international summit on AI safety. Seeing it happen with my own eyes – and done so surprisingly well – was very moving." Indeed, the symbolic significance of world leaders, including those from the US and China, sharing a stage and espousing a shared understanding of catastrophic risks represents a major milestone in international cooperation on AI.

# Futures

## Power of storytelling for social change

The Futures program aims to guide humanity towards the beneficial outcomes made possible by transformative technologies. By employing tools like storytelling, worldbuilding, scenario planning, research, and forecasting, we explore possible futures and determine actionable pathways. We then pinpoint the essential policies, decisions, and institutions needed to navigate these paths, and how best to deliver them. The program seeks to engage a diverse group of stakeholders from different professions, communities, and regions to shape our shared future together.



### Artificial Escalation

The Future of Life Institute partnered with Space Film & VFX on a short fictional film called “Artificial Escalation” which depicts a world where AI is integrated into nuclear command, control and communications systems (‘NC3’) with terrifying results. When disaster strikes, US, Chinese and Taiwanese military commanders quickly discover that with their new operating system in place, everything is sped up to unmanageable speeds. They have little time to work out what is going on, and even less time to prevent the situation escalating into a major catastrophe.



### 2023 Future of Life Award

The [Future of Life Award](#) honors those who have contributed to making our ongoing existence possible, yet have remained unacknowledged for their role in shaping a safer world. In an era shadowed by the Cold War’s persistent dread, two 1983 films resonated powerfully with the public and world leaders. The 2023 Future of Life Award honors the visionaries behind these narratives: Walter F. Parkes and Larry Lasker were behind the riveting techno-thriller “WarGames,” while the groundbreaking television event “The Day After” comes from the visionary executive Brandon Stoddard, with Edward Hume as screenwriter and Nicholas Meyer as director. The award, presented on the 40th anniversary of these films, recognizes the power of storytelling for creating a better, safer world. These films influenced policymakers and raised public awareness about the grave risks and apocalyptic consequences associated with nuclear warfare, leading to tangible, preventative action from leaders. Emilia Javorsky, Director of FLI’s Futures program explains: *“These films and their creators showcase the profound role that storytellers can play in tackling some of our world’s most intractable and extreme threats. They serve as a leading example of how artists can help make the world safer by examining urgent issues in compelling and evocative ways, and in turn, inspire our leaders to step up and take action. FLI is honored to celebrate them.”*

# Communications

FLI has continued to use communications and messaging to advocate for effective governance around transformative technologies, drive public education and awareness on the associated implications, risks and benefits, and highlight FLI's own activities. We work to build relationships across the media landscape in order to secure high-value coverage and create content in line with our organizational objectives. In the past six months alone, we have been mentioned in more than 1,300 unique articles. FLI staff engaged in more than 30 top-tier speaking engagements, including Reuters NEXT, Chatham House, Semafor and more.

Since the publication of the open letter in March 2023, we have continued to produce a drumbeat of communications to reiterate the need to effectively control and regulate the most advanced AI systems. We have worked with internal and external voices to produce content which advances our specific policy and reputational goals. Examples include:

- This [op-ed](#) in Le Monde, coauthored by FLI President Max Tegmark and The Elders Chair (and former Ireland President and UN Human Rights Commissioner Mary Robinson);
- This [interview](#) with Forbes feat. Max Tegmark, exploring the issue of deepfakes;
- This [interview](#) with The Guardian, talking about the need to regulate the out-of-control and irresponsible AI race to the bottom;

- This [video](#) produced by FLI to commemorate six months since we published the "Pause Letter";
- This Wired [article](#) on the growing global movement to ban lethal autonomous weapons, feat. FLI's Anna Hehir;
- This [op-ed](#) in The Hill by FLI Executive Director Anthony Aguirre about the urgent need to slow AI development and put necessary safeguards in place.

We collaborate with partner organizations such as Encode Justice, Control AI and The Elders with communications, utilizing our press networks and capabilities to place articles, secure interviews, and disseminate messaging to target audiences on key issues.

We have grown our podcast to reach new audiences (10,000+ active listeners), through guests such as AI ethics voice and youth activist Sneha Revanur. FLI staff appeared on other high-profile podcasts, including Axios' 1 Big Thing and Exponential. We increased our newsletter readership to 40,000.

We continue to produce homegrown materials and content to publish on our website and push across our networks, acting as a knowledge resource for those wishing to understand the implications and solutions around transformative technology.

# Grants

In 2023, we announced our second class of Vitalik Buterin Fellows in AI Existential Safety. We selected 3 new [PhD Fellows](#): Yawen Duan, Caspar Oesterheld, and Kayo Yin; and one new [Postdoctoral Fellow](#): Peter Park. We are committed to providing tuition, as well as a generous stipend and research fund throughout the Fellows' studies. In 2023, there were numerous opportunities for our Fellows to gather virtually, as well as in-person, along with other AI researchers committed to keeping AI safe and beneficial.

In 2023, we announced the winners of our first [RFP focused on the Humanitarian Impacts of Nuclear War](#). Over the next three years, we have committed millions of dollars of funding for scientific research on the horrific effects of nuclear conflict. Projects will focus on nuclear war scenarios, urban fires, crop production, agriculture and global food security, and effects on the ozone layer. The selected researchers are committed to collaboration and will meet throughout the grant period to share their work with each other and the broader research community.

KEY: *Grant Recipient / Amount Recommended (\$ USD)*

## **AI Impacts** \$162,000.00

General support. AI Impacts performs research related to the future of AI. They aim to answer decision-relevant questions in the most neglected areas of AI strategy and forecasting. The intended audience includes researchers doing work related to artificial intelligence, philanthropists involved in funding research related to artificial intelligence, and policy-makers whose decisions may be influenced by their expectations about artificial intelligence.

## **AI Objectives Institute** \$500,000.00

Support for the Talk to the City and Moral Mirror projects. AI Objectives Institute (AOI) is a non-profit research lab of leading builders and researchers. AOI brings together a network of volunteer researchers from top AI labs with product builders and experts from psychology, political theory and economics for a sociotechnical systems perspective on AI dynamics as well as post-singularity outcomes, improving odds that human values thrive in a world of rapidly deployed, extremely capable AI systems evolving with existing institutions and incentives.

## **Alignment Research Center** \$1,401,000.00

Support for the Alignment Research Center (ARC) Evaluation (Evals) Team. Evals is a new team at ARC building capability evaluations (and in the future, alignment evaluations) for advanced ML models. The goals of the project are to improve our understanding of what alignment danger is going to look like, understand how far away we are from dangerous AI, and create metrics that labs can make commitments around.

## **Berkeley Existential Risk Initiative** \$481,000.00

Support for the Berkeley Existential Risk Initiative's (BERI) collaboration with The Center for Human-Compatible Artificial Intelligence (CHAI). BERI's mission is to improve human civilization's long-term prospects for survival and flourishing. Currently, its main strategy is collaborating with university research groups working to reduce existential risk by providing them with free services and support. CHAI's mission is to develop the conceptual and technical wherewithal to reorient the general thrust of AI research towards provably beneficial systems.

## **Center for AI Safety** \$22,000.00

General support. The Center for AI Safety (CAIS) exists to ensure the safe development and deployment of AI. AI risk has emerged as a global priority, ranking alongside pandemics and nuclear war. Despite its importance, AI safety remains remarkably neglected, outpaced by the rapid rate of AI development. Currently, society is ill-prepared to manage the risks from AI. CAIS exists to equip policymakers, business leaders, and the broader world with the understanding and tools necessary to manage AI risk.

## **Center for Humane Technology** \$500,000.00

Support for AI-related policy work and messaging cohesion within the AI xrisk community. The Center for Humane Technology works to align technology with humanity's best interests. They envision a world with technology that respects our attention, improves our well-being, and strengthens communities.



**Centre for Long-Term Resilience (Alpenglow Group Limited) \$769,000.00**

General support. The Centre for Long-Term Resilience (CLTR) is an independent think tank based in the UK with a mission to transform global resilience to extreme risks. They do this by working with governments and other institutions to improve relevant governance, processes, and decision making. They believe that extreme risks pose some of the biggest challenges of the 21st Century, and want to work with a wide range of people and organisations who share their goals. As a result, they aim to collaborate with a broad coalition of policymakers, academics, think tanks and private-sector risk experts.

**Collective Intelligence Project \$150,000.00**

Support for the Alignment Assemblies for Legitimate AI Governance project. The Collective Intelligence Project is an experimental R&D organization that advances collective intelligence capabilities (meaning tools and systems for better constructing and cooperating towards shared goals) for effective collective governance of AI and other transformative technologies. Its mission is to direct technological development towards the collective good by making it possible to elicit and execute on collective values, and in doing so, to avoid disproportionate risks and harness benefits.

**FAR AI \$1,858,000.00**

General support. FAR accelerates neglected but high-potential AI safety research agendas. It supports projects that are either too large to be led by academia or overlooked by the commercial sector as they are unprofitable. FAR AI's mission is to ensure AI systems are trustworthy and beneficial to society.

**Foresight Institute \$290,000.00**

Support for the Existential Hope and Tech Tree Programs. Foresight Institute supports the development of beneficial high-impact technology to make great futures more likely. Their work to make flourishing futures for life more likely is mainly done by selectively advancing beneficial uses of high-impact science and technology via five technical interest groups focusing on 1) Molecular Machines: atomically precise control of matter, 2) Biotech: reverse aging and improve cognition, 3) Computer Science: secure decentralized human AI cooperation, 4) Neurotech: Improving Cognition, 5) Space: Expanding outward.

**The Future Society \$365,000.00**

General support. The Future Society (TFS) is an independent nonprofit organization based in US and Europe with a mission to align AI through better governance.

**Ought \$358,000.00**

General support. Ought is a product-driven research lab that develops mechanisms for delegating high-quality reasoning to advanced machine learning systems. Ought is building Elicit, a research assistant using language models to scale up high-quality reasoning in the world. Ought also conducts research to advocate for supervising the process of machine learning systems, not just their outcomes, so that we can avoid alignment risks from goal misspecification and opaque AI systems.

**Oxford China Policy Lab \$252,000.00**

General support. The Oxford China Policy Lab (OCPL) is a non-partisan group of researchers based at the University of Oxford. Its overarching goal is to mitigate the possibility of global risk associated with US-China great power competition (GPC), with a particular focus on risks stemming from AI and other emerging technologies.

**Redwood Research Group \$500,000.00**

General Support. Redwood Research Group's main work is doing theoretical and applied AI alignment research. They are especially interested in practical projects that are motivated by theoretical arguments for how the techniques we develop might successfully scale to the superhuman regime. They also run Constellation, a co-working space with members from ARC, MIRI, OpenPhil, Redwood, and some other organizations.

**University of California at Berkeley \$600,000.00**

The AI Policy Hub is housed at the AI Security Initiative at the University of California, Berkeley's Center for Long-Term Cybersecurity (CLTC), and the University of California's CITRIS Policy Lab at the Center for Information Technology Research in the Interest of Society and the Banatao Institute (CITRIS). The AI Policy Hub is training a new generation of AI policy talent to shape the future of artificial intelligence in the public interest. The Hub supports annual cohorts of graduate student researchers who conduct innovative research and make meaningful contributions to the AI policy landscape.

**The University of Chicago \$380,000.00**

Support for the University of Chicago's Existential Risk Laboratory (XLab). At this stage, XLab is focused on building student and faculty interest in existential risk through a postdoctoral fellowship, undergraduate courses, and a summer research fellowship. These activities are aimed at ultimately being able to offer a minor in Existential Risk Studies through an entity with active faculty research groups.

# Finances

FLI is a mission-driven organization and continually aims to diversify its funding. Shortly after we were founded in 2014, Elon Musk funded the world's first academic grant program in AI safety. Since those early days, we have received over 1,500 donations of various sizes from a wide variety of donors to fund our programs.

In 2021, computer programmer Vitalik Buterin provided FLI with a large and unconditional donation that in part serves as an endowment and helps to guarantee our independence.

## Expenditures for FLI activities 2023

Category	2023 Expenditure
Program: Grantmaking	\$9,936,998
Program: Scientific and Technical Education and Outreach Activities	\$3,199,855
Program: Research, Outreach and Advocacy regarding Policy	\$2,161,384
Management and General	\$1,019,475
<b>TOTAL 2023 Expenditure</b>	<b>\$16,317,712</b>

## Our income in 2023

FLI only engaged in limited fundraising in 2023 and our total income that year was about \$625,000. The difference between income and expenditure was covered out of our endowment.

# Team

## Board



**Anthony Aguirre,**  
Executive Director and  
Board Secretary



**Jaan Tallinn,**  
Board Member



**Max Tegmark,**  
President



**Meia Chita-Tegmark,**  
Board Treasurer



**Victoria Krakovna,**  
Board Member

## Staff

**Akhil Deo,** Social Media Manager

**Alexandra Tsalidis,** Policy Researcher

**Andrea Berman,** Grants Manager

**Anna Hehir,** Program Manager for  
Autonomous Weapons Systems

**Anna Yelizarova,** Special Projects Lead

**Ben Cumming,** Communications Director

**Ben Eisenpress,** Director of Operations

**Blake Pierson,** Director of Communications,  
Outreach & Operations

**Carlos Ignacio Gutierrez,** Policy Researcher

**David Nicholson,** Manager, Future of Life  
Award

**Emilia Javorsky,** Director Futures Program

**Gus Docker,** Podcast Host & Director

**Hamza Tariq Chaudhry,** U.S. Policy Specialist

**Landon Klein,** Director of U.S. Policy

**Maggie Munro,** Communications Specialist

**Mark Brakel,** Director of Policy

**Richard Mallah,** Principal AI Safety Strategist

**Risto Uuk,** EU Research Lead

**Taylor Jones,** Website Manager

**Tim Schreier,** Policy Researcher & Technical  
Standards Advisor

**Will Jones,** Associate, Futures Program

**General inquiries**

[contact@futureoflife.org](mailto:contact@futureoflife.org)

**Policy**

[policy@futureoflife.org](mailto:policy@futureoflife.org)

**Futures**

[futures@futureoflife.org](mailto:futures@futureoflife.org)

**Grants**

[grants@futureoflife.org](mailto:grants@futureoflife.org)

**Press**

[press@futureoflife.org](mailto:press@futureoflife.org)

