*Technology is giving life
the potential to flourish
like never before...*

*...or to self-destruct.
Let's make a difference!*
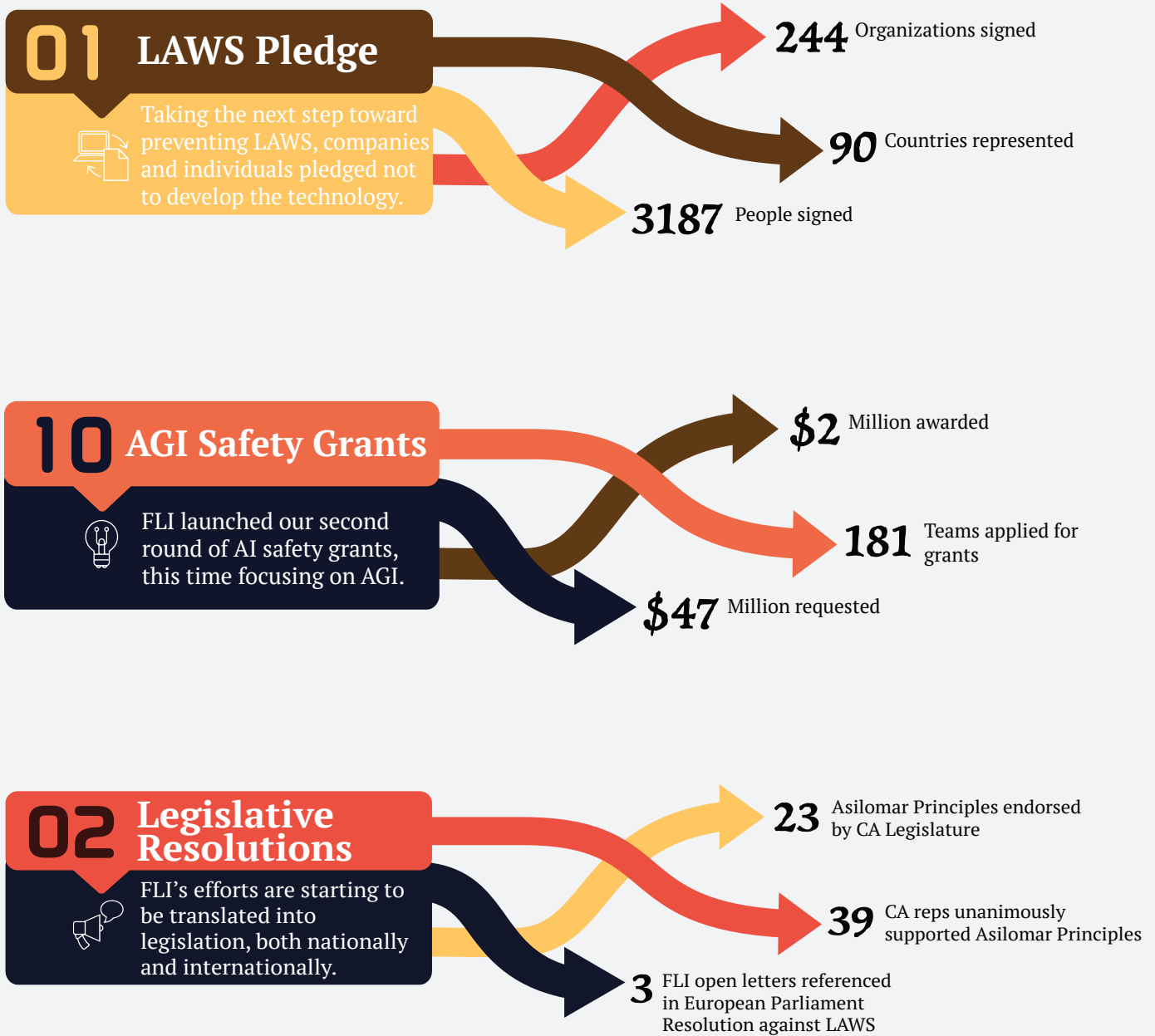
future of life

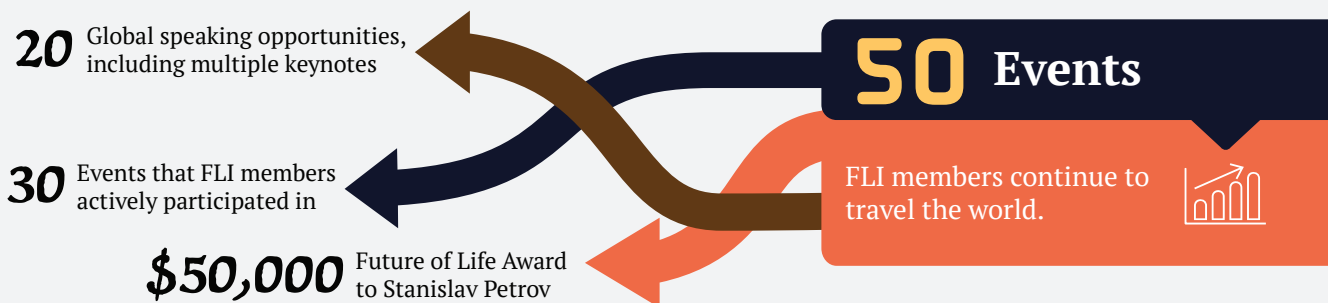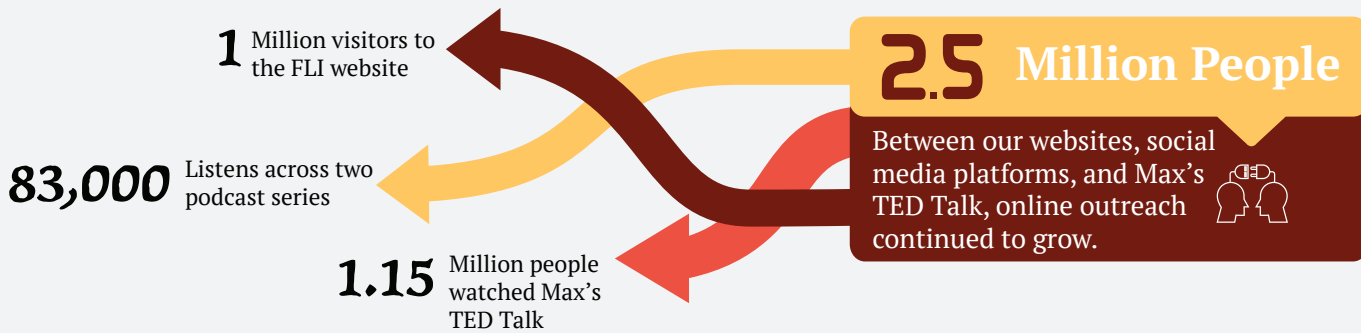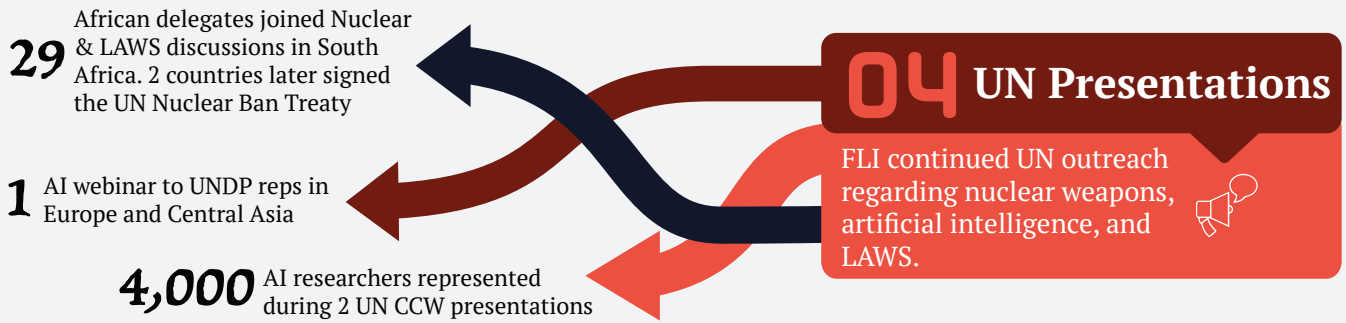## INSTITUTE

Annual Report
Making a Difference

2018

2018: By the Numbers

01 LAWS Pledge

Taking the next step toward preventing LAWS, companies and individuals pledged not to develop the technology.

244 Organizations signed

90 Countries represented

3187 People signed

10 AGI Safety Grants

FLI launched our second round of AI safety grants, this time focusing on AGI.

$2 Million awarded

181 Teams applied for grants

$47 Million requested

02 Legislative Resolutions

FLI's efforts are starting to be translated into legislation, both nationally and internationally.

23 Asilomar Principles endorsed by CA Legislature

39 CA reps unanimously supported Asilomar Principles

3 FLI open letters referenced in European Parliament Resolution against LAWS

29 African delegates joined Nuclear & LAWS discussions in South Africa. 2 countries later signed the UN Nuclear Ban Treaty

1 AI webinar to UNDP reps in Europe and Central Asia

4,000 AI researchers represented during 2 UN CCW presentations

04 UN Presentations

FLI continued UN outreach regarding nuclear weapons, artificial intelligence, and LAWS.

1 Million visitors to the FLI website

83,000 Listens across two podcast series

1.15 Million people watched Max's TED Talk

2.5 Million People

Between our websites, social media platforms, and Max's TED Talk, online outreach continued to grow.

20 Global speaking opportunities, including multiple keynotes

30 Events that FLI members actively participated in

$50,000 Future of Life Award to Stanislav Petrov

50 Events

FLI members continue to travel the world.

Message from the President

Will we humans wisely use our ever more powerful technology to end disease and poverty and create a truly inspiring future, or will we sloppily use it to drive ever more species extinct, including our own? We're rapidly approaching this fork in the road: the past year saw the power of our technology grow rapidly, with spectacular scientific breakthroughs from biotech to AI, while the wisdom with which we manage our technology remained far from spectacular: the UN CCW effort to limit lethal autonomous weapons collapsed, as did the Iran Deal and the landmark Reagan-Gorbachev INF nuclear disarmament treaty.

It's been a great honor for me to get to work with such a talented and idealistic team at our institute to ensure that tomorrow's technology helps humanity flourish rather than flounder. Thanks to Elon Musk, we were able to grant $2M to kickstart 10 projects around the world focused on keeping Artificial General Intelligence (AGI) safe and beneficial. We honored the man who helped avert WWIII in 1983 with the 2018 Future of Life Award, and we got leading AI companies and researchers to pledge not to build lethal autonomous weapons. The impact of the Asilomar AI Principles developed at our 2017 conference continued to grow, and they were officially endorsed by the State of California in August. Our largest undertaking of 2018 was to prepare the Puerto Beneficial AGI conference, which took place last month and was a great success; we look forward to building on many of the collaborations and initiatives that emerged from it for years to come.

In 2019, we look forward to results from the ongoing Beneficial-AGI grants program, to launching many of the initiatives that emerged from our Beneficial AGI conference, and to highlighting technological opportunities and challenges through education and outreach.

The best way for us to honor our founding advisory board member Stephen Hawking, who sadly left us last year, is to follow the advice he both gave and lived: even seemingly insurmountable challenges can be overcome with creativity, willpower and positive attitude. Technology is giving life the potential to flourish like never before, so let's seize this opportunity together!

Max Tegmark
February 2019

Major Accomplishments of 2018

Lethal Autonomous Weapons Systems (LAWS)

Following on the success of the Slaughterbots video at the end of 2017, which won a Gold Medal at the Cannes Corporate Media & TV Awards for best viral film of the year and now has an estimated 70 million views across various social media platforms, we continued working in support of a global ban on LAWS.

The LAWS Pledge

After years of voicing concerns, AI leaders have for the first time taken concrete action against lethal autonomous weapons, signing a pledge to "neither participate in nor support the development, manufacture, trade, or use of lethal autonomous weapons."

The pledge was announced by FLI president Max Tegmark at the annual International Joint Conference on Artificial Intelligence (IJCAI) in July, and was covered by CNN, The Guardian, NPR, Forbes, The Washington Post, and Newsweek. It has so far been signed by over 240 AI-related companies and organizations and more than 3,100 individuals from 90 countries. Signatories of the pledge include Google DeepMind, University College London, the XPRIZE Foundation, ClearPath Robotics/OTTO Motors, the European Association for AI (EurAI), the Swedish AI Society (SAIS), Demis Hassabis, British MP Alex Sobel, Elon Musk, Stuart Russell, Yoshua Bengio, Anca Dragan, and Toby Walsh.

The KAIST Boycott

After learning that the Korea Advanced Institute of Science and Technology (KAIST) planned to open an AI weapons lab in collaboration with a major arms company, AI researcher Toby Walsh led an academic boycott of the university. FLI assisted, and over 50 of the world's leading AI and robotics researchers from 30 countries joined the boycott. After less than a week, KAIST backtracked and agreed to "not conduct any research activities counter to human dignity including autonomous weapons lacking meaningful human control." The boycott was covered by CNN and The Guardian.

Working Toward a Global Ban on LAWS

In response to the pledge, FLI was invited by the Campaign to Stop Killer Robots to present at a side-event at the United Nations Convention on Conventional Weapons (UN CCW) meeting in August. FLI's Ariel Conn attended and gave a statement on behalf of both FLI and all of the scientists who have signed both the pledge and our other LAWS open letters. As a result, references to the Pledge and to FLI's previous LAWS open letters were included in the European Parliament's resolution calling on the United Nations to ban LAWS.

Also in August, Ariel helped organize a two-day, regional UN meeting for local African delegates to discuss issues pertaining to LAWS, as well as to discuss signing and ratifying the Treaty on the Prohibition of Nuclear Weapons. The meeting was held in Pretoria, South Africa and was attended by 29 delegates from 20 African countries, 2 of which subsequently signed the treaty.

AGI Safety Grants

AI safety is always our top priority, so we're excited to announce that we launched our second AI safety grant round in July. We awarded $2 million, donated by Elon Musk, to 10 grant winners to fund research aimed at ensuring artificial general intelligence (AGI), if it is ever developed, will be safe and beneficial.

Grant topics include: training multiple AI systems to work together and learn from humans, training AI to understand individual human preferences, quantifying what "general" actually means, incentivizing research groups to avoid a potentially dangerous AI race, and many more. As the request for proposals (RFP) stated, "The focus of this RFP is on technical research or other projects enabling development of AI that is beneficial to society and robust in the sense that the benefits have some guarantees: our AI systems must do what we want them to do."

FLI hopes that this round of grants will help ensure that AI remains beneficial as it becomes increasingly intelligent. The full list of FLI recipients and project titles can be found on page 13 of this document. This effort follows the first AI safety grant round in 2015, which funded 37 research teams. You can read more about the 2018 grant winners here.

CA State Legislation Supporting
23 Asilomar Principles

During 2018, FLI worked with California State Assemblyman Kevin Kiley to develop legislation in support of the 23 Asilomar AI Principles. On August 30, the California State Legislature unanimously adopted the new legislation in support of the Future of Life Institute's Asilomar AI Principles.

The Asilomar AI Principles are a set of 23 principles intended to promote the safe and beneficial development of artificial intelligence. The principles – which include research issues, ethics and values, and longer-term issues – have been endorsed by AI research leaders at Google DeepMind, GoogleBrain, Facebook, Apple, and OpenAI. Signatories include Demis Hassabis, Yoshua Bengio, Elon Musk, Ray Kurzweil, the late Stephen Hawking, Tasha McCauley, Joseph Gordon-Levitt, Jeff Dean, Tom Gruber, Anthony Romero, Stuart Russell, and more than 3,800 other AI researchers and experts.

With the new law (ACR 215) passing the State Senate with unanimous support, the California Legislature has now been added to that list.

2nd Future of Life Award Honors "The Man Who Saved the World"

To celebrate that September 26 was not the 35th anniversary of World War III, Stanislav Petrov, the man who helped avert an all-out nuclear exchange between Russia and the U.S., was honored with the $50,000 Future of Life Award at a ceremony at the Museum of Mathematics in New York City.

One of the closest calls in nuclear weapons history occurred thirty-five years ago, on September 26, 1983, when Petrov chose to ignore the Soviet early-warning detection system that had erroneously indicated five incoming American nuclear missiles. With his decision to ignore algorithms and instead follow his gut instinct, Petrov helped prevent an all-out US-Russian nuclear war, as detailed in the documentary film "The Man Who Saved the World", which was released earlier this year on Amazon, iTunes, and GooglePlay. Since Petrov passed away last year, the award was collected by his daughter Elena. In a symbolic manifestation of how frosty relationships have become between the two nuclear superpowers, Petrov's son could not attend because the US declined to approve his visa application.

The Future of Life Award seeks to recognize and reward those who take exceptional measures to safeguard the collective future of humanity. Last year's award was given to Vasili Arkhipov, who single-handedly prevented a nuclear attack on the US during the Cuban Missile Crisis. Petrov's Future of Life Award was covered in Vox, Daily Mail, Engineering 360, The Daily Star, and the Bulletin of Atomic Scientists.

Stanislav Petrov's daughter Elena holds the 2018 Future of Life Award flanked by her husband Victor. From left: Ariel Conn (FLI), Lucas Perry (FLI), Hannah Fry, Victor, Elena, Steven Mao (exec. producer of the Petrov film "The Man Who Saved the World"), Max Tegmark (FLI)

New Resource Pages

AI Policy Page

Artificial intelligence holds great economic, social, medical, security, and environmental promise. But in order to realize this potential, the challenges associated with AI development have to be addressed. This page highlights three complementary resources to help decision makers navigate AI policy: A global landscape of national and international AI strategies; a list of prominent AI policy challenges and key recommendations that have been made to address them; and a list of AI policy resources for those hoping to learn more.

Biotech Page

This page covers everything you need to know about biotechnology, including: what it is, a very brief history of the field, some of the latest breakthroughs and mishaps, the risks of unintended consequences, the risks of weaponizing biology, the ethics of biotechnology, and helpful overviews of the four main tools of biotech: DNA Sequencing, Recombinant DNA, DNA Synthesis, and Genome Editing. And for those who want to learn more, we've included a very extensive list of resources at the end of the page.
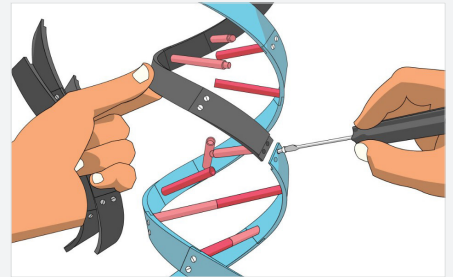
Nuclear Divestment Made Easy

Most money for nuclear weapons comes from taxes, but it's paid to companies that produce a variety of products. This page is dedicated to helping individuals, governments, and institutions reduce the risk from nuclear weapons by getting their money out of nuclear weapons production. Stigmatizing these companies with actions like divestment and boycotts can make it financially beneficial for them to pull out of nuclear weapons production — as occurred with landmines and cluster munitions.

Online Outreach

Our online outreach grew again this year, with the website drawing over 964,000 new visitors and well over a million visitors in total. The Benefits and Risks of Artificial Intelligence continued to be our most popular page by far, receiving around ¾ of a million readers. Our international outreach remained steady with about a third of our visitors coming from outside the United States, and about 250,000 people visited our translated pages.

FLI's podcast series have also grown. Ariel continued to produce a podcast each month featuring a discussion between experts on topics related to existential risks. Meanwhile, in April of this year, FLI's Lucas Perry launched a new podcast series on the AI value alignment problem. Lucas interviews technical and non-technical researchers in machine learning, AI safety, governance, coordination, ethics, philosophy, and psychology as they pertain to the project of creating beneficial AI. The two hosts produced 20 podcasts in 2018, which were listened to 83,000 times -- nearly triple the reach of podcasts in 2017.

Even more popular, by far, was Max's TED Talk, How to Get Empowered, Not Overpowered, by AI, which was released in June and has since been viewed 1.2 million times. In the talk, Max separates the real opportunities and threats of artificial intelligence from the myths, describing the concrete steps we should take today to ensure that AI ends up being the best — rather than the worst — thing to ever happen to humanity.

We're also pleased that our Slaughterbots video remained popular throughout 2018. It's been shared by dozens, if not hundreds, of organizations with an estimated 70 million views worldwide. We've also received over 60 media requests to share the video or clips from the video on news sites, in documentaries, in classrooms, and even as part of museum displays. We also had an interesting debate regarding the possibility of creating and defending against Slaughterbots with Paul Scharre, which can be found on the IEEE website.

FLI was mentioned in nearly every major US newspaper and tech publication this year, as well as many outside the US, including: CNN, Fortune, The Verge, Futurism, Forbes, Fox News, CNET, Gizmodo, The Independent, Quartz, Defense One, the Washington Post, Newsweek, ZDNet, the New York Post, Bloomberg, the New Yorker, and many more.

Volunteer of the Year

Na Li (Lina) has been volunteering with the Future of Life Institute since June of 2016. Her helpful and effective personality led her to being involved with a number of writing and research projects during her time here. Primarily, she has been instrumental in setting up the Chinese translations now available on our website. In addition to writing these translations herself, Lina has helped to build a small team of volunteers within FLI dedicated to this undertaking. Under her mentorship, this team has contributed over 30 Chinese translations to the site, and Chinese now ranks as the 4th most common language spoken by people visiting the site.

AGI Safety Grant Recipients

Primary Investigator	Project Title	Amount
Allan Dafoe, Yale University	Governance of AI Programme	$276,000
Stefano Ermon, Stanford University	Value Alignment and Multi-agent Inverse Reinforcement Learning	$100,000
Owain Evans, Oxford University	Factored Cognition: Amplifying Human Cognition for Safely Scalable AGI	$225,000
The Anh Han, Teesside University	Incentives for Safety Agreement Compliance in AI Race	$224,747
Jose Hernandez-Orallo, University of Cambridge	Paradigms of Artificial General Intelligence and Their Associated Risks	$220,000
Marcus Hutter, Australian National University	The Control Problem for Universal AI: A Formal Investigation	$276,000
James Miller, Smith College	Utility Functions: A Guide for Artificial General Intelligence Theorists	$78,289
Dorsa Sadigh, Stanford University	Safe Learning and Verification of Human-AI Systems	$250,000
Peter Stone, University of Texas	Ad hoc Teamwork and Moral Feedback as a Framework for Safe Robot Behavior	$200,000
Josh Tenenbaum, MIT	Reverse Engineering Fair Cooperation	$150,000

In 2018, we spent $3M, and most of it went to research...

12% AI Conferences & Workshops

3% Nuclear Safety Projects

2% Project Support

8% Outreach & Education

75% Beneficial AI Research Projects
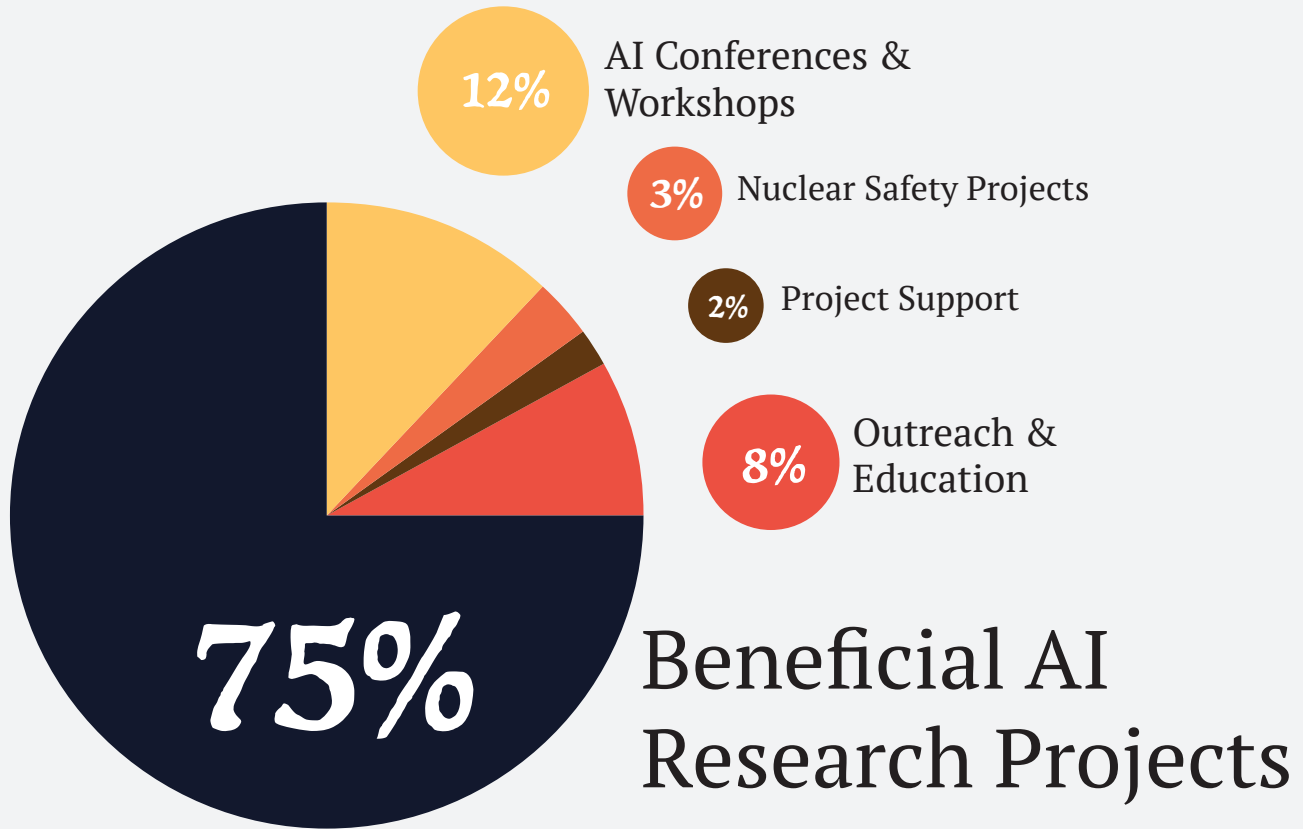
We at FLI are proud that almost all of our income goes straight to our programs: in 2018, we spent $0 on fundraising and very little on administration.

Most of the $3 million spent in 2018 went to the 10 new research teams around the world to whom we awarded grants for keeping AGI beneficial, though some of it continued to support AI safety research from our first round of grants. Our outreach spending was primarily directed to our website (futureoflife.org) and toward continued efforts to support scientists and AI industry leaders as they call on the United Nations to begin negotiations to ban lethal autonomous weapons. Our conference spending was mainly for the Beneficial AGI meeting held in Puerto Rico at the start of 2019. Our nuclear safety projects included the second $50k Future of Life Award for Stanislav Petrov, who helped prevent WWIII.

Thanks!

We would like to extend a special thank you to the donors who made all these great accomplishments possible: 100% of the $3M described on the previous page came from philanthropic donations. We are especially grateful to Elon Musk for his generous donation, which enabled the continuation of our beneficial AI grants program.

We are also deeply thankful to the Open Philanthropy Project, the Berkeley Existential Risk Initiative, Jaan Tallinn, George Godula, Alexander Tamas, and Lee Oxman, and the many other donors whose generous support helped make possible everything we have done so far.

Jaan Tallinn

Open Philanthropy Project

Berkeley Existential Risk Initiative

And, of course, we wouldn't be nearly as successful without the help of our volunteers!
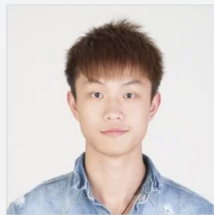
Jacob Beebe	Yishuai Du	Olivier Grondin	You Jeen Ha	Anna Karpunina
Na Li (Lina)	Roman Maas	Jason Orlosky	Grzegorz Orwinski	Kevin Wang	Kristy Wen
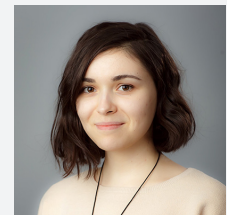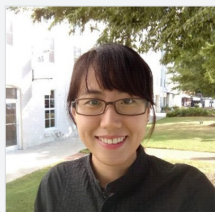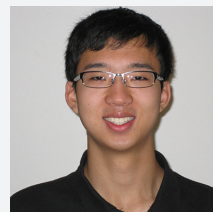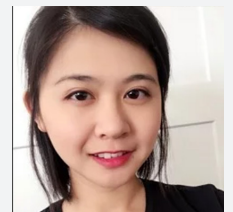
Team

FOUNDERS

Jaan Tallinn Max Tegmark Meia Chita-Tegmark Viktoriya Krakovna Anthony Aguirre
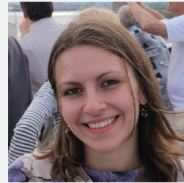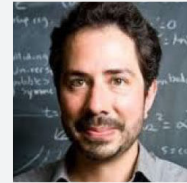
SCIENTIFIC ADVISORY BOARD
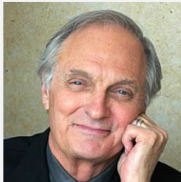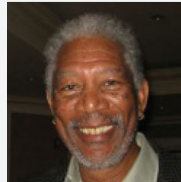
Alan Alda Nick Boström Erik Brynjolfsson George Church Morgan Freeman Alan Guth
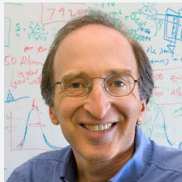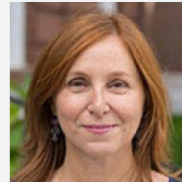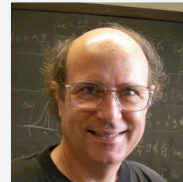
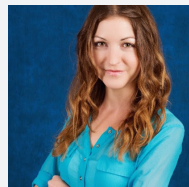Christof Koch Elon Musk Saul Perlmutter Martin Rees Francesca Rossi Stuart Russell Frank Wilczek

We at FLI are deeply saddened to note the passing of Scientific Advisory Board member Stephen Hawking in March, 2018.
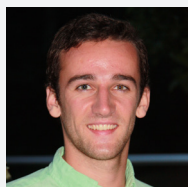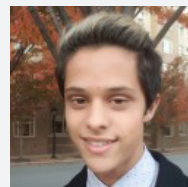
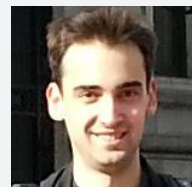## CORE TEAM

Ariel Conn Jessica Cussins Tucker Davey Richard Mallah Lucas Perry David Stanley

Memoriam to Stephen Hawking

by Max Tegmark

As we mourn the loss of Stephen Hawking, we should remember that his legacy goes far beyond science. Yes, of course he was one of the greatest scientists of the past century, discovering that black holes evaporate and helping found the modern quest for quantum gravity. But he also had a remarkable legacy as a social activist, who looked far beyond the next election cycle and used his powerful voice to bring out the best in us all. As a founding member of FLI's Scientific Advisory board, he tirelessly helped us highlight the importance of long-term thinking and ensuring that we use technology to help humanity flourish rather than flounder. I marveled at how he could sometimes answer my emails faster than my grad students. His activism revealed the same visionary fearlessness as his scientific and personal life: he saw further ahead than most of those around him and wasn't afraid of controversially sounding the alarm about humanity's sloppy handling of powerful technology, from nuclear weapons to AI.

On a personal note, I'm saddened to have lost not only a long-time collaborator but, above all, a great inspiration, always reminding me of how seemingly insurmountable challenges can be overcome with creativity, willpower and positive attitude. Thanks Stephen for inspiring us all!

future of life INSTITUTE

npr AI Innovators Take Pledge Against Autonomous Killer Weapons

Let's Talk About AI Ethics; We're On A Deadline **Forbes**

IEEE SPECTRUM Debating Slaughterbots and the Future of Autonomous Weapons

HOW FRIGHTENED SHOULD WE BE OF A.I.? THE NEW YORKER

Vox 35 years ago today, one man saved us from world-ending nuclear war

The Dawn of Killer Robots **VICE**

The Atlantic **Why Earth's History Appears So Miraculous**

Scientists call for boycott of South Korean university over killer robot fears **CNN**

The Guardian Thousands of leading AI researchers sign pledge against killer robots

Making a Difference