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

...or to self-destruct. Let's make a difference!

Annual Report
Making a Difference

2017
2017 By the Numbers

**01 Beneficial AI Conference**
Our sequel to the Puerto Rico Conference produced the Asilomar AI Principles.
- 146 Participants
- 23 Asilomar AI Principles
- 1273 AI researcher signatories

**01 UN Nuclear Ban**
The United Nations adopted a historic treaty banning nuclear weapons.
- 3754 Scientists signed an open letter presented to the UN in support of a nuke ban
- 31 Nobel laureates were among our open letter signatories
- 15 Signatories of the open letter participated in a video presented to the UN

**46 Videos Posted to YouTube**
We reached tens of millions of people this year thanks to videos that went viral.
- 1 Slaughterbots video seen by tens of millions
- 6 Videos with over 10,000 views
- 63 Million views worldwide
Million people visited FLI and autonomousweapons.org

1.2

02 Websites
Our new website about lethal autonomous weapons helped us double our reach in 2017.

Million visits to the Benefits & Risks of AI page

.6

50,000 Followers across social media channels

Websites

Articles published worldwide that reference FLI

550

12 FLI in the News
We were mentioned in major news outlets every month.

Articles published about Slaughterbots

100

113 Articles written for the FLI site

FLI in the News

Total number of donations

160

Donors
We’ve achieved so much thanks to the generosity of our donors.

4 Donors who gave at least $100,000

101

8 Donors who gave at least $5,000

Donors

Articles written for the FLI site
Message from the President

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 most powerful technologies help humanity flourish rather than flounder.

A prime example of poor technology use is a global system where nuclear war occurs through miscalculation, error, or terror if we merely wait long enough. 2017 saw a growing specter of nuclear war, fueled by souring rhetoric and a new nuclear arms race between the US, Russia, and North Korea. On an optimistic note, 2017 also saw growing awareness and stigmatization of this development thanks to the new UN nuclear weapons ban and ensuing Nobel Peace Prize; we worked hard to support such stigmatization by delivering an open letter from thousands of scientists to the UN and honoring the man who prevented WWIII in 1962 with a well-publicized award.

We similarly worked to stigmatize destabilizing weaponization of AI with an open letter from industry leaders and a viral video.

After helping mainstream AI safety with our 2015 Puerto Rico conference and subsequent research grant program, we are focusing our efforts on the topic that remains most controversial – and ultimately most important: what will happen if Artificial General Intelligence (AGI) outperforms humans on all cognitive tasks within decades, as most AI-researchers expect. Our 2017 Asilomar conference was a milestone in mainstreaming AGI safety: Whereas the open letter that emerged from our 2015 Puerto Rico conference merely spoke in vague terms about the importance of keeping AI beneficial, the Asilomar AI Principles that emerged from our 2017 conference had real teeth: they explicitly mention recursive self-improvement, superintelligence and existential risk, yet are signed by a veritable who’s who of leading AI researchers.

In 2018, we look forward to running a new grant program adding momentum to AGI safety, as well as to highlighting technological opportunities and challenges through education and outreach. Technology is giving life the potential to flourish like never before – let’s seize this opportunity together!

-Max Tegmark
Major Accomplishments of 2017

Participants at the 2017 Beneficial AI Conference in Asilomar, CA.

Beneficial AI

To kick off 2017, FLI organized and hosted our biggest event yet: the Beneficial AI (BAI) Conference and workshop in Asilomar, California. This five-day event in January followed our 2015 Puerto Rico conference, but it had even more high-level participants, and it advanced the AI safety agenda significantly further.

The event began with a two-day workshop for FLI’s 37 AI grant recipients to highlight and discuss their research progress. In 2017 alone, our grantees produced 33 academic papers, and participated in 118 conferences and workshops.

The workshop was immediately followed by the main 2.5-day conference. The 146 attendees included AI researchers Stuart Russell, Yoshua Bengio, and Francesca Rossi, economists Jeffrey Sachs and Erik Brynjolfsson, philosopher Nick Bostrom, ACLU president Anthony Romero, psychologist Daniel Kahneman, and industry leaders Elon Musk, Larry Page, and Eric Schmidt. The participants hashed out opportunities and challenges related to the future of AI and considered steps we can take to ensure that the technology remains beneficial. Videos of the talks and discussion panels can be found on the FLI YouTube channel, including the panel discussion on superintelligence, which received over 300K views.
Beneficial AI: The Asilomar AI Principles

A key goal of the conference was to formulate a set of principles that AI researchers, social scientists, industry leaders, and laypeople could agree on as we move closer to advanced AI. After months of research and discussions with AI thought leaders, FLI distilled a list of principles that were vigorously discussed and edited by the Asilomar attendees. This produced the 23 Asilomar AI Principles, which have now been signed by over 1,200 AI researchers from around the world, including AI leaders from DeepMind and other top technology companies. The principles have been mentioned in major news outlets including Inverse, Newsweek, Business Insider, and The Guardian.

Asilomar participants discussing details about what would become the 23 Asilomar Principles.

The Asilomar AI Principles cover both near-term issues such as research culture, privacy, transparency and value alignment, as well as longer-term issues related to superintelligence. Whereas the open letter that emerged from our 2015 Puerto Rico conference (and helped mainstream AI safety) merely spoke in vague terms about the importance of keeping AI beneficial, the Asilomar Principles had real teeth: they explicitly mention recursive self-improvement, superintelligence and existential risk.

To continue the discussion, we began a series of articles about the Principles, tackling one Principle at a time, and encouraging readers to consider how these Principles can be implemented in society. We look forward to expanding this conversation in 2018 and working directly with organizations and researchers to ensure AI advances in accordance with the Asilomar AI Principles.
Beneficial AI: New AI Safety Grants

As 2017 began with AI safety, so it ended: we’re excited to have launched a second AI safety grants competition in December. With near-term AI safety research now more firmly established, this second round focuses on safety issues associated with Artificial General Intelligence (AGI): how can we ensure that AI remains safe and beneficial even as it approaches and surpasses human level on all cognitive tasks?

UN Ban Treaty: Stigmatizing the New Nuclear Arms Race

In October of 2016, the UN General Assembly voted to begin negotiations on a treaty to ban nuclear weapons, which started in March of 2017 and concluded in July. Any nation adopting the ban treaty agrees never to develop or host nuclear weapons.

FLI is in favor of a future of life where zero nuclear wars have occurred, and views this UN initiative as valuable for stigmatizing excessive nuclear arsenals: the superpowers have way more nuclear weapons than required for deterrence, and instead of abiding by their disarmament pledges from the non-proliferation treaty, they are developing new ones that are destabilizing and more “usable” while proliferation continues elsewhere.
Stigmatizing the New Nuclear Arms Race, Cont.

In early 2017, FLI began collaborating with the International Campaign to Abolish Nuclear Weapons (ICAN), which had led the efforts behind the negotiations. We first released an open letter from scientists in support of the UN negotiations. Over 3,700 scientists from over 100 countries signed the letter, including 31 Nobel Laureates, Stephen Hawking, and former US Secretary of Defense, William Perry. FLI presented the letter and signatures on a poster to the UN delegates during the March negotiations, as featured in the New York Times.

In June, during the second round of negotiations, FLI created and presented a five-minute video featuring statements from Nobel Laureates and nuclear experts to the delegates at the United Nations, urging them to support the ban treaty.

The UN negotiations ended in July with the adoption of a treaty that categorically bans all nuclear weapons. The “Treaty on the Prohibition of Nuclear Weapons” was adopted with 122 votes in favor, one vote against, and one country abstaining. So far, 53 countries have signed it and 3 have ratified it.

ICAN, led by Beatrice Fihn (left), was awarded the Nobel Peace Prize for their incredible work galvanizing international support for the ban.

In 2018, we look forward to collaborating more with ICAN as they work to encourage more countries to sign the treaty and continue their efforts to reduce the threat of nuclear weapons.

In addition to our work with ICAN and the UN, FLI co-organized and participated in a conference at MIT this May, titled “Reducing the Threat of Nuclear Weapons.” Speakers included California Congresswoman Barbara Lee, former Secretary of Energy, Ernest Moniz, and Joe Cirincione, President of the Ploughshares Fund.
Slaughterbots: Stigmatizing Lethal Autonomous Weapons

Support for an international ban on lethal autonomous weapons systems has been mounting inside and outside the AI community for the past few years. In 2015, FLI partnered with AI researchers Stuart Russell and Toby Walsh on an open letter opposing offensive autonomous weapons, which was signed by thousands of AI researchers and over 20,000 people in all. We supported Walsh on another open letter this year, from leaders of major artificial intelligence companies and directed at the United Nations Convention on Conventional Weapons (UN CCW), which was scheduled to discuss LAWS in November of this year. The letter was signed by over 130 leaders of AI companies in 26 countries, including Elon Musk (Tesla, SpaceX, OpenAI) and Demis Hassabis and Mustafa Suleyman (Google’s DeepMind).

To help people understand the issue, FLI teamed up with AI pioneer Stuart Russell to create a short film illustrating the unique terror that autonomous weapons systems could bring upon the world.

The seven-minute film, “Slaughterbots,” depicts a disturbing future where lethal autonomous weapons have become cheap, ubiquitous, and extremely difficult to stop. Russell presented it at a UN CCW side event in November, which was hosted by the Campaign to Stop Killer Robots. In an appearance at the end of the video, Russell warns that the technology described in the film already exists at a basic level, and that the window to act before it gets integrated, miniaturized and mass-produced as in the film is closing fast.

Within days, Slaughterbots went viral and has now received around 60 million views across different platforms. Major news outlets picked it up, as well, including CNN, The Washington Post, VICE, Fox News, and The Guardian.
“Ban Lethal Autonomous Weapons”

Coinciding with the November launch of Slaughterbots, we built a website, autonomousweapons.org, that provides educational information about lethal autonomous weapons, the organizations who are working to ban them, and what concerned citizens can do. Over 200,000 people visited the site during the month and a half it was live in 2017. We look forward to growing our audience more in 2018, and we hope this momentum will encourage more policymakers and industry leaders to take the issue seriously.

The Inaugural Future of Life Award

This October, FLI presented our inaugural Future of Life Award, an award dedicated to someone who has done an amazing service to humanity and didn’t receive proper recognition at the time. This year’s recipient was the soft-spoken naval officer, Vasili Arkhipov (left), who arguably saved the world by single-handedly preventing a Soviet nuclear attack during the height of the Cuban Missile Crisis. Arkhipov’s submarine captain, thinking their vessel was under attack by American forces, wanted to launch a nuclear weapon at the ships above, which would probably have ignited World War III. But Arkhipov, with the power of veto, said no. It is sobering that very few have heard of Arkhipov, although his action might be considered the most valuable individual contribution to human survival in modern history.

Arkhipov’s daughter Elena and grandson Sergei (right) flew from Moscow to London for the award ceremony at the Institute of Engineering & Technology. After explaining Arkhipov’s heroics to the audience, FLI president Max Tegmark presented the Arkhipov family with their award and $50,000. The event was covered by The Times, The Guardian, The Independent, and The Atlantic.
In August, Max Tegmark released his new book, Life 3.0: Being Human in the Age of Artificial Intelligence. In it, he explores how AI will impact life as it grows increasingly advanced, perhaps even achieving superintelligence. While many books on artificial intelligence focus on the near-term impacts of AI on jobs and the military, they often avoid critical longer-term questions: Will superhuman artificial intelligence arrive in our lifetime? Can and should it be controlled, and if so, by whom? Can humanity survive in the age of AI? And if so, how can we find meaning and purpose if super-intelligent machines provide for all our needs?

Tegmark’s book directly focuses on these questions, so that everyone can join what he considers to be “the greatest conversation of our time.” Life 3.0 received positive reviews from Ray Kurzweil, Stuart Russell, and Yuval Noah Harari, as well as in Nature, Science, The Times, The Guardian and other prominent papers. It became a New York Times best seller, an International Best Seller and an Amazon Top 10. Tegmark was also interviewed on NPR, ScienceFriday, and Sam Harris’s Waking Up podcast to discuss the book. FLI posted surveys from Life 3.0 on our website and received high engagement from our readers. Tegmark distilled core ideas from the book into a MinutePhysics video about superintelligence, which has over 700,000 views. We’re excited to continue building on the success of Max’s book in 2018 and to broaden the conversation even more.
Online Outreach

2017 saw continued growth for the FLI website and for our national and international profile, as nearly a million people from around the world visited our website. Our most popular page, the Benefits and Risks of Artificial Intelligence, was read by over half a million people. Meanwhile, over 47,000 people read our most popular article of the year, “90% of All the Scientists That Ever Lived Are Alive Today,” by Eric Gastfriend, while tens of thousands of others read the various discussions about the Asilomar AI Principles.

The FLI podcast has been listened to by tens of thousands of people, and our website's interactive features continued to remain popular. The new Superintelligence Survey, which went along with Max’s Life 3.0, drew over 50,000 people, and our older nuclear apps continued to draw people by the tens of thousands. Additionally, we’ve translated 55 of our most popular pages into 9 different languages, helping draw over a quarter of a million non-English speakers to the site.


Most Read Articles of 2017

1.) 90% of All the Scientists That Ever Lived Are Live Today
   By Eric Gastfriend

2.) AI Researchers Create Video to Call for Autonomous Weapons Ban at United Nations
   By Jessica Cussins

3.) A Principled AI Discussion in Asilomar
   By the FLI Team

4.) Developing Countries Can’t Afford Climate Change
   By Tucker Davey

5.) How Do We Align Artificial Intelligence With Human Values?
   By Ariel Conn
Volunteers of the Year

Eric Gastfriend has been with FLI since the first days of its founding, and he’s been heading FLI’s Japan outreach and translation efforts, leading a team of volunteers fluent in both Japanese and English. Over the past year, Eric and his team have fostered relations with Dwango AI Lab and the Whole Brain Architecture Initiative (WBAI), both based in Japan and focused on AI R&D. Through these collaborations, they published a Japanese translation of FLI’s research priorities document in the Japanese Society for Artificial Intelligence Journal. They also interviewed Hiroshi Yamakawa, the Chief of Dwango AI Lab and chairperson at WBAI, discussing the state of AI research in Japan and its societal implications in comparison to those in the western world. Eric also wrote this year’s most popular article on FLI, “90% of All the Scientists That Ever Lived Are Alive Today,” which as mentioned above has been read by over 47,000 people to date.

Jacob Beebe joined FLI a year and a half ago and has been one of our most active volunteers. In particular, he is heavily involved with our Russian outreach efforts, having co-managed the recruitment and interviewing of new volunteer translators, and also proofreading their translations. Additionally, Jake has helped with FLI’s nuclear weapons awareness projects, conducting research into the social and geopolitical reasons for why countries hold a particular stance on nuclear weapons. He is currently also our “czar” of Russian social media, having set up and managed FLI’s profiles there. We look forward to working with him more in the new year.
In 2017, we spent $2M, and most of it went to research...

We at FLI are proud that almost all of our income goes straight to our programs: in 2017, we spent $0 on fundraising and very little on administration.

Most of the 2 million dollars spent in 2017 went to the 37 research teams around the world to whom we awarded grants for keeping AI beneficial: most of these payments are second installments of three-year projects. Our outreach spending was mainly on our website (futureoflife.org) and educational videos, and our conference spending was mainly for the Asilomar AI meeting. Our nuclear safety projects included the $50k Future of Life Award for the submariner who prevented WWIII.
2017 AI Safety Publications

Alex Aiken

Peter Asaro

Vincent Conitzer
- Conitzer, V. et al. Moral Decision Making Frameworks for Artificial Intelligence. AAAI 2017, Duke University, Durham, NC 27708, USA
- Deng, Y. and Conitzer V. Disarmament Games. AAAI 2017, Duke University Durham, NC 27708, USA

Stefano Ermon

Owain Evans
Fuxin Li

Percy Liang

Long Ouyang

David Parkes
- Calibrated fairness in bandits (Christos Dimitrakakis, Yang Liu, Debmalya Mandal, David Parkes, and Goran Radanovic), In Fairness, Accountability and Transparency in Machine Learning Workshop, at KDD, 2017.

Andre Platzer

Francesca Rossi
Stuart Russell

Bas Steunebrink

Michael Wellman

Brian Ziebart
Thank You!

We would like to extend a special thank you to the donors who made all these great accomplishments possible: 100% of the $2M 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, Matt Wage, Lee Oxman, Brian Corcoran, Aaron Merchak, and the many other donors whose generous support helped make possible everything we have done so far.

And, of course, we wouldn’t be nearly as successful without the help of our volunteers.
Meet the 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  Stephen Hawking  Christof Koch  Elon Musk  Saul Perlmutter  Martin Rees  Francesca Rossi  Stuart Russell  Frank Wilczek

Core Team

Ariel Conn  Jessica Cussins  Tucker Davey  Richard Mallah  Lucas Perry  David Stanley
'Slaughterbots' film shows potential horrors of killer drones

UC Berkeley professor’s ‘slaughterbots’ video on killer drones goes viral

A physicist explores the future of artificial intelligence

When the World Lucked Out of a Nuclear War

Military robots are getting smaller and more capable
Soon, they will travel in swarms

“...scientists bear a special responsibility for nuclear weapons...”

The Economist

The Times

Soviet officer Vasili Arkhipov, who averted nuclear war, is honoured

Futurism

World’s Top Experts Have Created a “Law of Robotics”

Artificial intelligence: how scared should we be about machines taking over?

The Telegraph

Making a Difference