

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



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

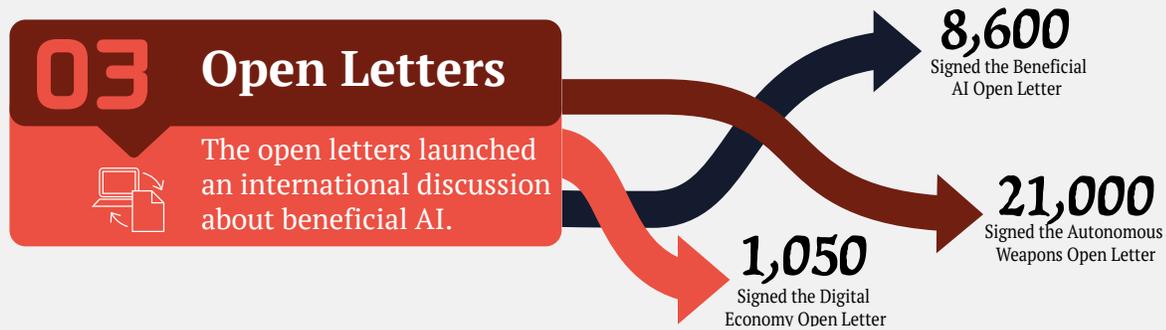
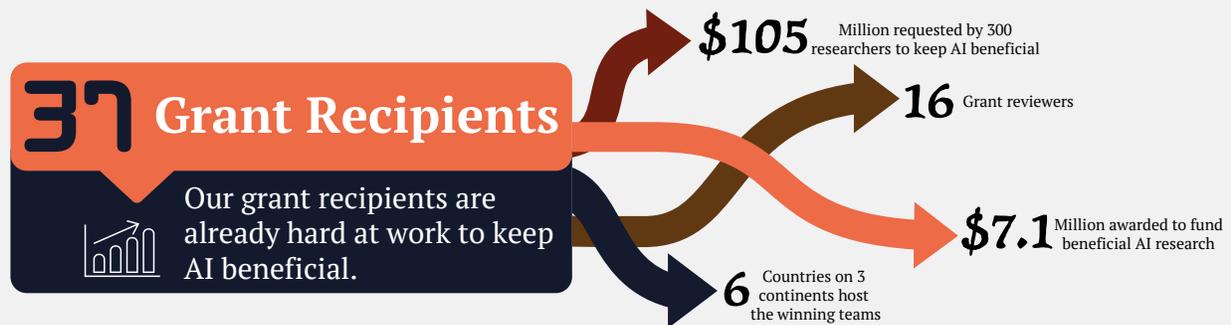
future  
of life  
INSTITUTE

# Annual Report

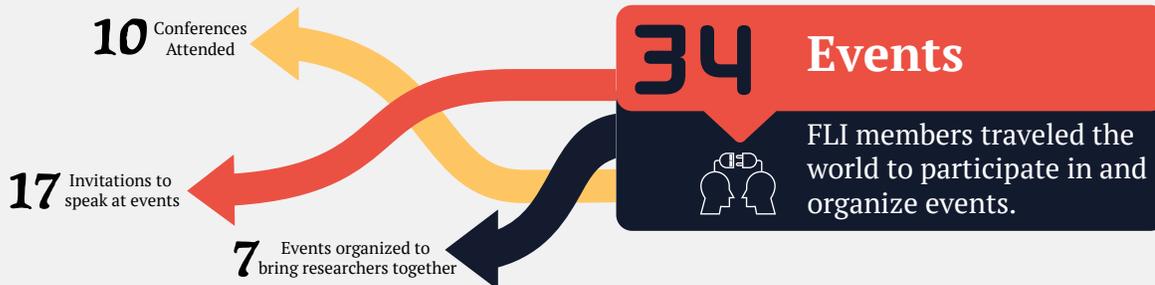
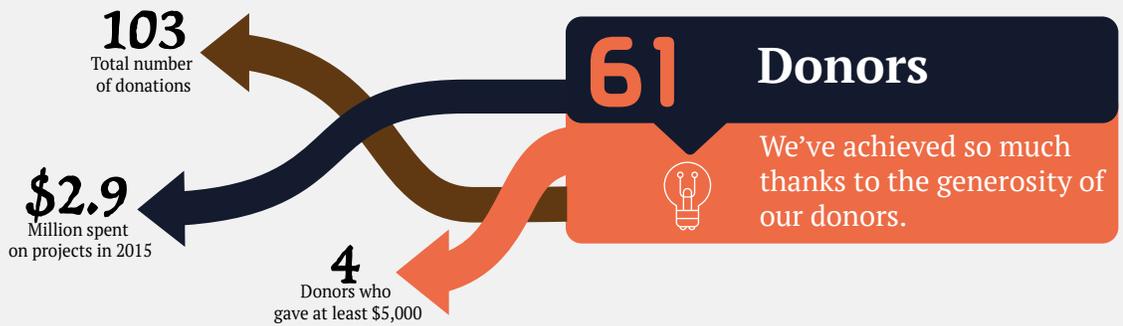
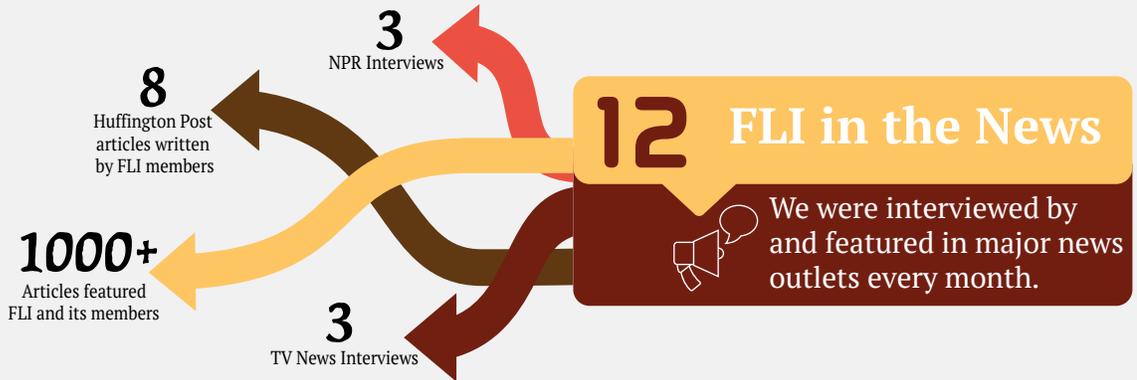
*Making a Difference*

2015

# 2015 By the Numbers



# 2015 By the Numbers, Cont.



## Message from the President

We have technology to thank for all the ways in which today is better than the stone age, and technology is improving at an accelerating pace. 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 are beneficial for humanity. With less powerful technologies such as fire, we learned to minimize risks largely by learning from mistakes. With more powerful technologies such as nuclear weapons, synthetic biology and future strong artificial intelligence, planning ahead is a better strategy than learning from mistakes. We've worked hard in 2015, supporting research and other efforts aimed at avoiding problems in the first place.

We are delighted to have helped the beneficial AI movement go mainstream in 2015 and to have launched the first-ever global research program aimed at keeping AI beneficial, awarding \$7.1M to 37 talented teams. Our freshly launched news site, [futureoflife.org](http://futureoflife.org), is reaching ever more people, and we're excited to have begun our first explorations into reducing risks from nuclear weapons and biotechnology.

Our institute brings together ideas, people and resources to make the future of life better. The above-mentioned ideas are brought to fruition by our remarkably talented and idealistic team, and since we're almost all volunteers working for free, our visionary donors can see their investments efficiently converted into impact, not administration.

As we look ahead to 2016 and beyond, we see great opportunities to build on and expand our beneficial AI efforts, to develop a similar program for beneficial biotechnology, to reduce the risk of accidental nuclear war, and to scale up our fledgling outreach efforts to give the quest for the long-term success of life the prominence it deserves. Technology is giving life the potential to flourish like never before – let's seize this opportunity together!

-Max Tegmark



# Major Accomplishments of 2015



*The Puerto Rico conference brought together some of the greatest minds in artificial intelligence research to start considering how we can keep AI beneficial. This one conference helped launch all of the great work we were able to accomplish this year.*

## Puerto Rico Conference

At the end of 2015, a Washington Post article described 2015 as the year the beneficial AI movement went mainstream. Only 360 days earlier, at the very start of the year, we were hosting the inaugural Puerto Rico Conference that helped launch that mainstreaming process. The goal of the conference was to bring more attention to the problems advanced artificial intelligence might bring and to galvanize the world's talent to solve them.

The three big results of the conference were:

- 1) Worldwide attention to our Beneficial AI Open Letter, which garnered over 8600 signatures from some of the most influential AI researchers, some of the most influential scientists in other fields, and thousands of other AI researchers and scientists from around the world.
- 2) The creation of our Research Priorities Document, which became the basis for our beneficial AI research grant program.
- 3) Elon Musk's announcement of a \$10 million donation to help fund the grants.

## Research to keep AI beneficial

With our goal now established, we launched a search for grant recipients. Over 300 PIs and research groups applied for funding to be part of the new field of AI safety research. On July 1, we announced the 37 AI safety research teams who would be awarded a total of \$7 million for the first round of research. With their projects now underway, we look forward to tracking their progress in the coming year.



*Our main objective for 2015 was to bring about useful discussion and research around how to keep AI safe and beneficial, and the main results were our beneficial AI grants program and our open letters.*

## More Open Letters

The success of the Beneficial AI open letter to build support for the beneficial AI movement prompted researchers to work with FLI to write two more open letters:

1) The Autonomous Weapons Open Letter - FLI worked with Stuart Russell, Toby Walsh and others to launch an open letter calling for a ban on offensive autonomous weapons. Walsh and Russell presented the letter at the International Joint Conference on Artificial Intelligence, making headlines around the world. By the end of the year, it had received over 22,000 signatures from scientists and citizens internationally, including over 3,000 AI and robotics researchers and 6 past presidents of the Association for the Advancement of Artificial Intelligence.

2) The Open Letter on the Digital Economy - Inspired by the success of the AI Open Letter, Erik Brynjolfsson, a member of the scientific advisory board, spearheaded an effort to bring attention to the rising inequality that technological advances could exacerbate. He and a team of economists and business leaders wrote an open letter about AI's future impact on the economy, which has received over 1,000 signatures to date, including 5 Nobel Laureates.

## Outreach

To broaden our quest to keep technology beneficial, we reached out to researchers in various fields as well as the general public. We hosted many small events to bring researchers together, we spoke at numerous conferences, symposia and workshops, and we established partnerships that will help broaden our scope of work and the audience we can reach. We also launched our news site, [futureoflife.org](http://futureoflife.org), to add detailed and nuanced reporting on the most powerful technologies that shape our future.

Although our biggest success stories of 2015 were in the AI area, we also laid the foundation for new projects in other fields in 2016. Among other things, we established working relationships with various biotech researchers, which will help us expand our research goals and outreach into promoting safe and beneficial biotech research. We also began exploring projects for reducing the risk of accidental nuclear war.



*FLI's Richard Mallah speaking at the NIPS symposium, *Algorithms Among Us: The Societal Impacts of Machine Learning*, which he also helped organize.*



*This year's Nobel Week Dialogue was on The Future of Intelligence, and FLI's Stuart Russell and Max Tegmark were invited as speakers to discuss how we can keep AI safe and beneficial.*

## The Beneficial AI Movement Is Growing

As a result of our Puerto Rico conference, the Open Letters and the research grants, many other groups, including the mainstream AI community picked up on the importance of the beneficial AI movement. The first annual International Workshop on AI, Ethics, and Society was held at AAAI in spring of 2015. FLI's Marc Warner worked with CSER to host an event at the Policy Exchange in London. Max Tegmark spoke at both the United Nations and the Nobel Week Dialogue about the future of AI. FLI's Richard Mallah helped organize a popular symposium at NIPS about keeping AI beneficial, and he started an FLI-IEEE partnership. FLI's advisor Stephen Hawking explained the goals of the beneficial AI movement at a Google conference, and he participated in a popular Reddit AMA about the impact of artificial intelligence. FLI's work was also featured or mentioned in well over 1000 news articles and reports from CBS, CNN, FOX, NBC, NPR, BBC, New York Times, Washington Post, Huffington Post, Wired, Popular Science, CNET, Financial Times, The Independent, The Telegraph, Live Science, and many other media outlets.

## Reducing the Risk of Nuclear War

Toward the end of the year, we began exploratory work aimed at reducing the risk of accidental nuclear war, an oft-ignored risk likely to kill more people than nuclear terrorism on average. FLI's nuclear goals are quite moderate: we simply oppose human extinction and other global catastrophes jeopardizing our civilization. We therefore support significantly reducing today's nuclear arsenals, ending hair-trigger alert, and stigmatizing the development and production of new nuclear weapons. Our initial focus is on empowering individuals and organizations to make a difference. For example, many invest (without even knowing it) their own money in producing new nuclear weapons systems that arguably make their country less safe. We're developing tools to make it easy you to find out and, if you so wish, to find alternative mutual funds that don't fuel the nuclear arms race.

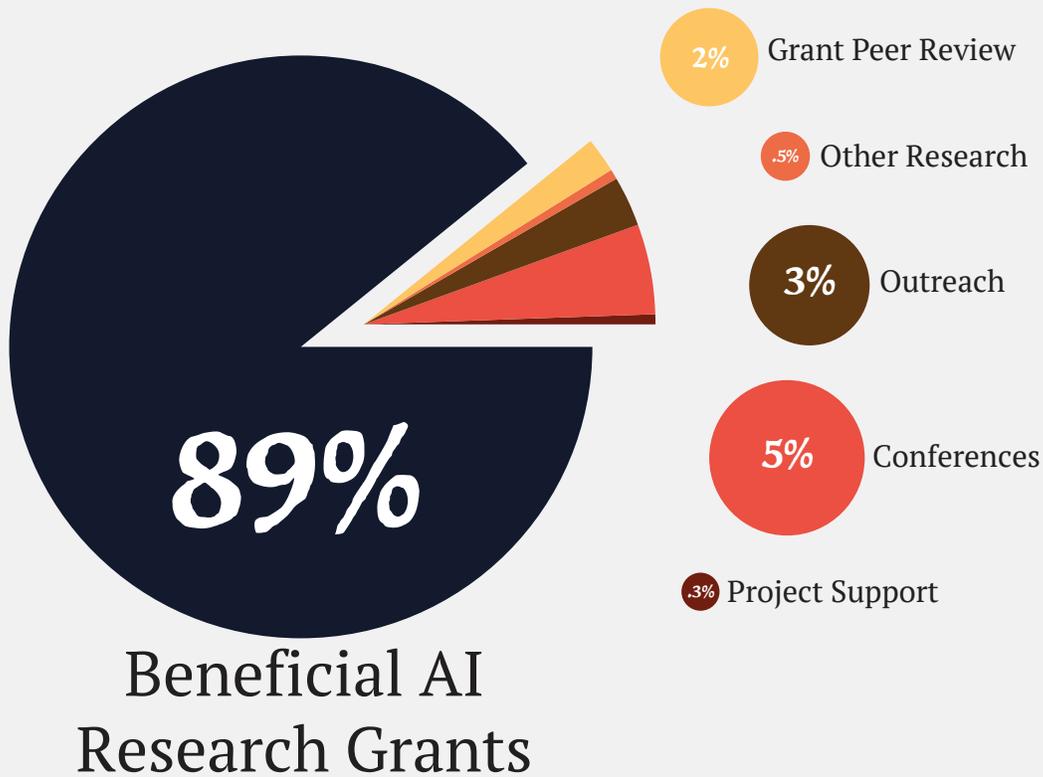


### Volunteer Highlight: Richard Mallah



One of FLI's hardest working volunteers is Richard Mallah. While working a full-time job for Cambridge Semantics, Richard has found time to attend nearly a dozen AI conferences and workshops around the world. He helped organize some of these events, and he was a panelist or speaker at even more of them. Over the year, he's helped establish some of our most important partnerships and he brought more attention to the open letters, garnering thousands of supporters. He also wrote FLI's most popular article to date, "The Top AI Breakthroughs of 2015."

In 2015, we spent \$2.9M, and most of it went to research...



We at FLI are proud that almost all our income goes straight to our programs: in 2015, we spent \$0 on fundraising and almost nothing on administration. This efficiency is possible because our leadership and administration is mainly done by our unpaid board of directors and volunteers, so that we incur costs only for tasks we can't do ourselves (such as filing legal documents).

Most of the 2.9 million dollars spent in 2015 went to the 37 research teams around the world to whom we awarded grants for keeping AI beneficial: most of these payments are first installments to multi-year projects which, if successful, will continue to receive funding for 1-2 more years. Our outreach spending (3%) was mainly on our news site <http://futureoflife.org> and our conference spending (5%) was mainly on our Puerto Rico meeting.

## 2015 Project Grants Recommended for Funding

<i>Primary Investigator</i>	<i>Project Title</i>	<i>Amount Recommended</i>
Alex Aiken, Stanford University	Verifying Deep Mathematical Properties of AI Systems	\$100,813
Peter Asaro, The New School	Regulating Autonomous Artificial Agents: A Systematic Approach to Developing AI & Robot Policy	\$116,974
Seth Baum, Social & Environmental Entrepreneurs	Evaluation of Safe Development Pathways for Artificial Superintelligence	\$100,000
Paul Christiano, University of California, Berkeley	Counterfactual Human Oversight	\$50,000
Vincent Conitzer, Duke University	How to Build Ethics into Robust Artificial Intelligence	\$200,000
Owen Cotton-Barratt, Centre for Effective Altruism, Oxford	Decision-relevant uncertainty in AI safety	\$119,670
Thomas Dietterich, Oregon State University	Robust and Transparent Artificial Intelligence Via Anomaly Detection and Explanation	\$200,000
Stefano Ermon, Stanford University	Robust probabilistic inference engines for autonomous agents	\$250,000
Owain Evans, University of Oxford	Inferring Human Values: Learning “Ought”, not “Is”	\$227,212
Benja Fallenstein, Machine Intelligence Research Institute	Aligning Superintelligence With Human Interests	\$250,000
Katja Grace, Machine Intelligence Research Institute	AI Impacts	\$49,310
Seth Herd, University of Colorado	Stability of Neuromorphic Motivational Systems	\$98,400

Ramana Kumar, University of Cambridge	Applying Formal Verification to Reflective Reasoning	\$36,750
Fuxin Li, Georgia Institute of Technology	Understanding when a deep network is going to be wrong	\$121,642
Percy Liang, Stanford University	Predictable AI via Failure Detection and Robustness	\$255,160
Long Ouyang, Theiss Research	Democratizing Programming: Synthesizing Valid Programs with Recursive Bayesian Inference	\$99,750
David Parkes, Harvard University	Mechanism Design for AI Architectures	\$200,000
Andre Platzer, Carnegie Mellon University	Faster Verification of AI-based Cyber-physical Systems	\$200,000
Heather Roff, University of Denver	Lethal Autonomous Weapons, Artificial Intelligence and Meaningful Human Control	\$136,918
Francesca Rossi, University of Padova	Safety Constraints and Ethical Principles in Collective Decision Making Systems	\$275,000
Benjamin Rubinstein, The University of Melbourne	Security Evaluation of Machine Learning Systems	\$98,532
Stuart Russell, University of California, Berkeley	Value Alignment and Moral Metareasoning	\$342,727
Bart Selman, Cornell University	Scaling-up AI Systems: Insights From Computational Complexity	\$24,950
Kaj Sotala, Theiss Research	Teaching AI Systems Human Values Through Human-Like Concept Learning	\$20,000
Bas Steunebrink, IDSIA	Experience-based AI (EXPAI)	\$196,650
Manuela Veloso, Carnegie Mellon University	Explanations for Complex AI Systems	\$200,000
Michael Webb, Stanford University	Optimal Transition to the AI Economy	\$76,318

Daniel Weld, University of Washington	Computational Ethics for Probabilistic Planning	\$200,000
Adrian Weller, University of Cambridge	Investigation of Self-Policing AI Agents	\$50,000
Michael Wellman, University of Michigan	Understanding and Mitigating AI Threats to the Financial System	\$200,000
Michael Wooldridge, University of Oxford	Towards a Code of Ethics for AI Research	\$125,000
Brian Ziebart, University of Illinois at Chicago	Towards Safer Inductive Learning	\$134,247

### 2015 Center Grant Recommendation

<i>Primary Investigator</i>	<i>Project Title</i>	<i>Amount Recommended</i>
Nick Bostrom, University of Oxford	Strategic Research Center for Artificial Intelligence	\$1,500,000

### 2015 Conference and Education Grant Recommendations

<i>Primary Investigator</i>	<i>Project Title</i>	<i>Amount Recommended</i>
Wendell Wallach, Yale	Control and Responsible Innovation in the Development of Autonomous Machines	\$180,000
Anna Salamon, Center for Applied Rationality	Specialized rationality skills for the AI research community	\$111,757
Moshe Vardi, Rice University	Artificial Intelligence and the Future of Work	\$69,000
Jacob Steinhardt, Stanford University	Summer Program in Applied Rationality and Cognition	\$88,050

# Thank You!

We would like to extend a special thank you to the donors who made all these great accomplishments possible: 100% of the \$2.9M described on the previous page came from philanthropic donations. We are especially grateful to Elon Musk for his generous donation, which enabled the creation of our beneficial AI grants program. We are also deeply thankful to the Open Philanthropy Project, whose generous support enabled this grant program to reach its full potential.



In addition, we wish to thank our co-founder Jaan Tallinn, who single handedly funded the Puerto Rico conference, as well as Matt Wage, Nisan Stiennon, Sam Harris, and Jacob Trefethen who helped make possible everything else that we have done so far.



Jaan Tallinn



Matt Wage



Nisan Stiennon



Sam Harris



Jacob Trefethen

We are also grateful to the Silicon Valley Community foundation for their wonderful help in managing the donor advised fund that has enabled our grantmaking.



An additional thank you must also go out to our smaller donors and our volunteers who made FLI a priority this year. Without your help, we could never have accomplished as much as we did.

## 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 & Volunteers



Ariel Conn



Jesse Galef



Richard Mallah



Lucas Perry



David Stanley



James Babcock



Riley Carney



Ales Flidr



Eric Gastfriend



Peter Haas



Max Kesin



Janos Kramar



Daniel Miller



Chase Moores

# THE NEW YORKER

DECEMBER 30, 2015

## THE BEST QUOTES FROM THE TECH INDUSTRY IN 2015

1. "The key question for humanity today is whether to start a global AI arms race or to prevent it from starting."

—An [open letter](#) from Elon Musk, Stephen Hawking, and many others about the dangers of artificial intelligence.

**theguardian**

Experts including Elon Musk call for research to avoid AI 'pitfalls'

An open letter from AI researchers warns of pitfalls ahead, and lays out a plan for avoiding them while improving the quality of artificial intelligence



**INDEPENDENT**

**Stephen Hawking, Elon Musk and others call for research to avoid dangers of artificial intelligence**

Document also signed by prominent employees of companies involved in AI, including Google, DeepMind and Vicarious



TECH & SCIENCE

**HAWKING, MUSK, WOZNIAK: BAN ARTIFICIALLY INTELLIGENT, AUTONOMOUS WEAPONS IN WAR**

More than 1,000 artificial intelligence experts, including Stephen Hawking, Elon Musk and Steve Wozniak, have [signed an open letter](#) calling for the ban of fully autonomous weaponry —weapons capable of killing without operators.

**Elon Musk-backed group gives \$7M to explore artificial intelligence risks**

The Future of Life Institute, which aims to protect humanity from the possible negative effects of artificial intelligence, awards the cash to 37 research teams.

**c|net**

**The New York Times**

***Elon Musk and Stephen Hawking Among Hundreds to Urge Ban on Military Robots***

**The Washington Post**

Elon Musk, Stephen Hawking, Google researchers join forces to avoid 'pitfalls' of artificial intelligence

**WSJ. D**

**Musk, Hawking Warn of Artificial Intelligence Weapons**

*Making a Difference*