



The story of FLI

# Future of Life Institute (FLI):

**Mission:** *FLI catalyzes and supports research and initiatives for safeguarding life and developing optimistic visions of the future, including positive ways for humanity to steer its own course considering new technologies and challenges.*

## Founders:



Jaan Tallinn



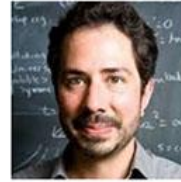
Max Tegmark



Meia Chita-Tegmark

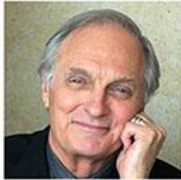


Viktoriya Krakovna



Anthony Aguirre

## Scientific Advisory Board:



Alan Alda



Nick Boström



Erik Brynjolfsson



George Church



Morgan Freeman



Alan Guth



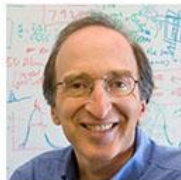
Stephen Hawking



Christof Koch



Elon Musk



Saul Perlmutter



Martin Rees



Francesca Rossi



Stuart Russell



Frank Wilczek

# How FLI formed

**Aug 2011:** Jaan meets Max and Anthony at FQXi conference and argues for AI risk

**Oct 2013:** Jaan visits Boston LessWrong house and introduces me to Max

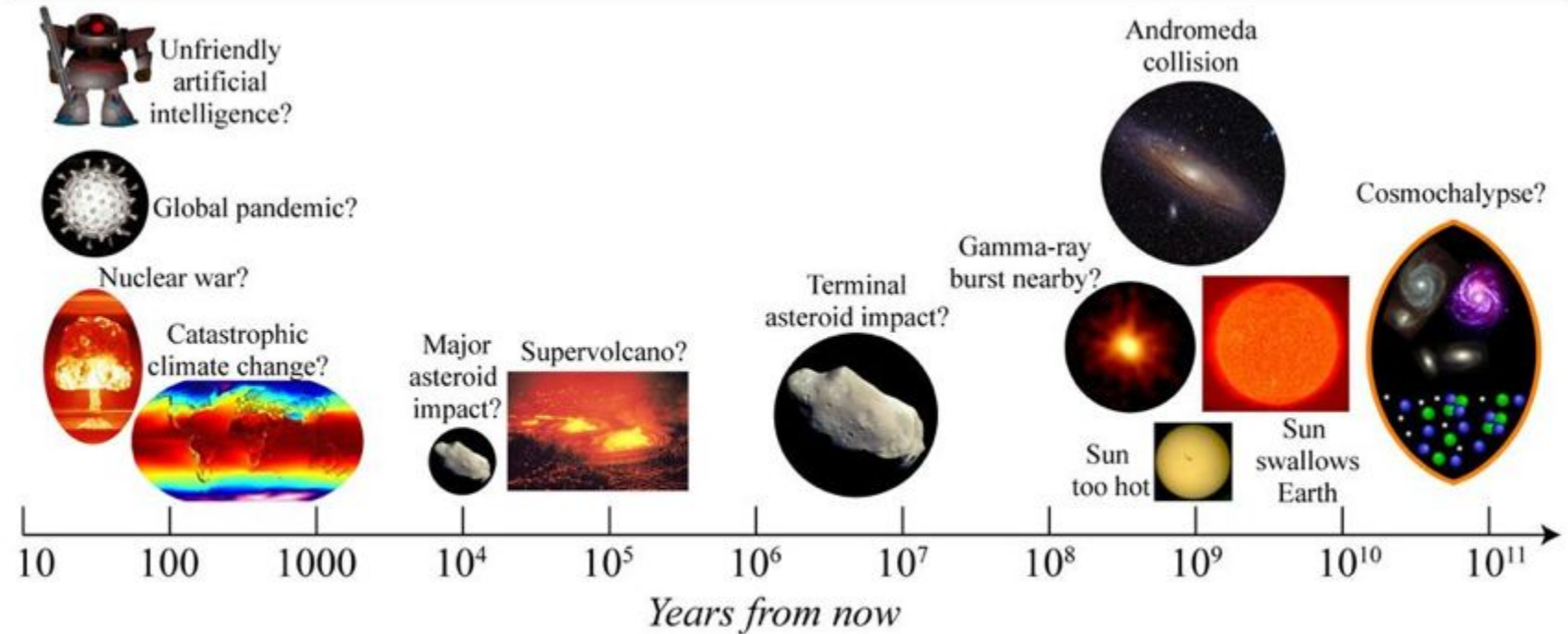
**Dec 2013:** Max and Meia decide to start x-risk organization modeled after FQXi, and get Anthony and Jaan on board

**Feb 2014:** Max and Meia invite me to join and organize volunteer efforts

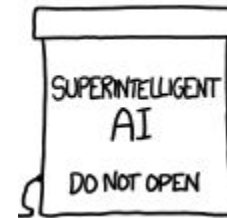
**March 2014:** First FLI meeting with local scientists and volunteers



# Existential risks



# Focus on AI safety



- At the **first meeting**, many were concerned about long-term AI risk
- Decided to **organize a conference on AI safety** to connect mainstream AI researchers with AI risk proponents
- Compiled a document of **research priorities** for the conference



# AI safety conference in January 2015

- Attended by top AI researchers (academia and industry)
- Experts endorsed open letter on research priorities



# Open letter on robust & beneficial AI

To date, the open letter has been signed by over 8,600 people. The list of signatories includes:

**Stuart Russell**, Berkeley, Professor of Computer Science, director of the Center for Intelligent Systems, and co-author of the standard textbook Artificial Intelligence: a Modern Approach.

**Tom Dietterich**, Oregon State, President of AAAI, Professor and Director of Intelligent Systems

**Eric Horvitz**, Microsoft research director, ex AAAI president, co-chair of the AAAI presidential panel on long-term AI futures

**Bart Selman**, Cornell, Professor of Computer Science, co-chair of the AAAI presidential panel on long-term AI futures

**Francesca Rossi**, Padova & Harvard, Professor of Computer Science, IJCAI President and Co-chair of AAAI committee on impact of AI and Ethical Issues

**Demis Hassabis**, co-founder of DeepMind

**Shane Legg**, co-founder of DeepMind

**Mustafa Suleyman**, co-founder of DeepMind

**Dileep George**, co-founder of Vicarious

**Scott Phoenix**, co-founder of Vicarious

**Yann LeCun**, head of Facebook's Artificial Intelligence Laboratory

**Geoffrey Hinton**, University of Toronto and Google Inc.

**Yoshua Bengio**, Université de Montréal

**Peter Norvig**, Director of research at Google and co-author of the standard textbook Artificial Intelligence: a Modern Approach

**Oren Etzioni**, CEO of Allen Inst. for AI

**Guruduth Banavar**, VP, Cognitive Computing, IBM Research

**Michael Wooldridge**, Oxford, Head of Dept. of Computer Science, Chair of European Coordinating Committee for Artificial Intelligence

**Leslie Pack Kaelbling**, MIT, Professor of Computer Science and Engineering, founder of the Journal of Machine Learning Research

**Tom Mitchell**, CMU, former President of AAAI, chair of Machine Learning Department

# Research grants program

**Jan 2015:** Announced 3-year international grants program to fund AI safety research (**\$10M funding from Elon Musk**)

**Mar 2015:** Received **~300 applications** totalling \$100M in requested funds

**Jul 2015:** **Awarded grants to 37 teams:** 32 project grants, 1 center grant, 4 outreach grants

**Sep 2015:** Research projects started.

## TECHNOLOGY

**Elon Musk donated \$10 million to ensure robots don't eventually murder you**

NICK WHIGHAM NEWS.COM.AU JANUARY 16, 2015 3:40PM

SHARE [f](#) [t](#) [in](#) [g+](#) [✉](#)

[SAVE THIS STORY](#)



Elon Musk has likened a potential worst case scenario of AI to the terminator films.



# Events & outreach



**Dec 2015:** co-organized **NIPS** symposium on societal impacts of machine learning

**Feb 2016:** Co-organized **AAAI** workshop on AI, Ethics & Society: FLI grant winners presenting

**Jul 2016:** Sponsored ICML workshop on Reliable ML

**Dec 2016:** Sponsoring NIPS workshop on Intepretable ML

Ongoing informal outreach to AI researchers at conferences

# Priorities going forward

- Continue **supporting research and discussion** on keeping AI beneficial
- Continue bringing AI safety **into the mainstream**
- Organize more **technical workshops**
- Apply lessons from our AI successes to **other risk areas**
- Continue **growing our news site** [futureoflife.org](https://futureoflife.org) to provide quality reporting on powerful emerging technologies

