SINGULARITY INSTITUTE

STRATEGIC PLAN

UPDATED: AUGUST 2011

SINGULARITY INSTITUTE - LETTER FROM THE PRESIDENT

Since 2000, Singularity Institute has been a leader in studying the impact of advanced artificial intelligence on the future. We have expanded to become a networking center for scientists, entrepreneurs, and investors involved in emerging technologies. Our annual conference, the Singularity Summit, attracts about 1,000 attendees, with coverage from top media such as *TIME*, *NPR*, *Scientific American*, *Forbes*, *Wired*, and *The New York Times*.

Every year, machines surpass human ability in new domains. When they surpass our ability to design artificial intelligence, they will be able to recursively improve their own intelligence in a way that could quickly lead to vastly more intelligent agents — an 'intelligence explosion'. If these agents are programmed with stable, beneficial goals, their arrival could bring unprecedented well-being to Earth and beyond.

Writing about the growing Singularity movement in a 2010 New York Times op-ed, cyber-space pioneer Jaron Lanier wrote,

> "These are ideas with tremendous currency in Silicon Valley; these are guiding principles, not just amusements, for many of the most influential technologists."

Singularity Institute exists to ensure that the powerful effects of advanced machine intelligence are beneficial for society. Our research and academic outreach has set the agenda for this important project concerning the future of humanity, and we see many opportunities in the future for additional progress.

Singularity Institute's Strategic Plan outlines the organization's goals and near-term strategies for achieving them. Our plan seeks to advance cutting-edge research, cultivate a broad community that can help us achieve our mission, and bring improvements in organizational transparency and effectiveness. This plan will help ensure that our leadership in planning for advanced artificial intelligence continues into the next decade and beyond.

Sincerely,

Michael Vassar
President, Singularity Institute

Our Mission

To ensure that the creation of smarter-than-human intelligence benefits society.

Our Values

Safety. Just as engineers build extra safety measures into the design of a bridge, we believe it is important to build extraordinary safety measures into smarter-than-human intelligence because it is an engineering project that may impact billions of lives.

Rationality. Rationality is a toolkit built from cognitive science, probability theory, and decision theory. We believe that the methods of rationality are useful for solving difficult problems like those of advanced artificial intelligence.

Empiricism. We believe that we must test our beliefs against reality in order to have an accurate model of the world and its opportunities.

Precision. We believe that the most important problems can only be solved by picking the exactly correct answer out of a large space of apparently possible answers.

Our History

Singularity Institute was founded in 2000 by artificial intelligence researcher Eliezer Yudkowsky and internet entrepreneurs Brian and Sabine Atkins. Today, it is recognized as a leading organization dedicated to researching artificial intelligence risks and planning for positive results from the development of advanced artificial intelligence.

Key achievements include:

- **Foundational publications on AI risks and Friendly AI**: *Basic AI Drives and Catastrophic Risks* (2010), *Timeless Decision Theory* (2010), *Convergence of Expected Utility for Universal Artificial Intelligence* (2009), *Artificial Intelligence as a Positive and Negative Factor in Global Risk* (2008), *Coherent Extrapolated Volition* (2004), *Creating Friendly AI* (2001), and others.

- **The Singularity Summit**: This annual TED-style conference features speakers who are leaders in business, science, and technology. The event highlights the opportunities and risks inherent in the future development of advanced technology and focuses the brightest minds on confronting these challenges.

- **LessWrong.com**: This online community with over 100,000 unique visitors a month studies the subjects of rationality, probability theory, decision theory, cognitive science, and catastrophic risks. It has produced an expanding network of local chapters that act as forums for rationality training, discussion, study, research, and self-improvement in London, New York, Paris, Los Angeles, San Francisco, Washington DC, and over 20 other locales worldwide.

Current Situation

Singularity Institute is currently the only organization dedicated to mitigating catastrophic risks from artificial intelligence. We have a stock of knowledge and analysis about AI risks accumulated through years of dedicated thinking on the topic. We also sit at the center of a large network of people interested in reducing AI risks.

The Singularity Summit is growing as a source of funding. It now enjoys a track record of prestigious speakers (e.g. Ray Kurzweil, Peter Norvig, Douglas Hofstadter, John Tooby, Stephen Wolfram), which makes it easier to recruit popular speakers and sustain the event. Some Singularity Summit speakers have gone on to produce important work relevant to our core mission, including David Chalmers, Steve Omohundro, and James Miller.

We face opportunities to (1) become more transparent to donors and collaborators, (2) spin off rationality training to another organization so that Singularity Institute can focus on Friendly AI research, and (3) build synergy with the optimal philanthropy, critical thinking, and humanism communities, which tend to share many of our values and goals.

Our Core Strategies

To achieve our core mission and "ensure that the creation of smarter-than-human intelligence benefits society," we have three core strategies:

1. Solve open research problems in Friendly AI theory and other fields relevant to our core mission.
2. Mobilize additional human capital and other resources for achieving our mission.
3. Improve the function and capabilities of the organization.

Strategy #1: **Solve open research problems in Friendly AI theory and other fields relevant to our core mission.**

1. Clarify the open problems relevant to our core mission.
2. Identify and recruit researcher candidates who can solve research problems.

    a. Visit universities to find potential researchers.
    b. Build relationships with professors who are teaching relevant materials to potential research candidates.
    c. Encourage university professors with access to large numbers of relevant students to draw attention to key ideas.
    d. Visit Top Coder, Math Olympiad, and similar events to find potential researcher candidates.
    e. Recruit more research associates and hire more researchers.
3. Use researchers and research associates to solve open problems related to Friendly AI theory.
4. Estimate current AI risk levels.
    a. Estimate prospects for AGI development.
    b. Create detailed models of pathways to intelligence explosion and the causal factors involved.

Strategy #2: Mobilize additional human capital and other resources for achieving our mission.

1. Continue operation of the Singularity Summit, which is beginning to yield a profit while also reaching more people with our message.
    a. Gain corporate sponsors for the Singularity Summit.
    b. Improve and expand the Singularity Summit outside the USA (e.g. Australia), if these events can be run profitably.
2. Cultivate LessWrong.com and the greater rationality community as a resource for Singularity Institute.
    a. Recruit supporters via LessWrong.com.
    b. Make use of LessWrong.com for collaborative problem-solving (in the manner of the earlier LessWrong.com progress on decision theory).
    c. Keep LessWrong.com active by publishing top-level posts.
    d. Create materials that will help Less Wrong meetups to be more fun and more productive.
    e. Encourage improvements in critical thinking in the wider world. We need a larger community of critical thinkers for use in recruiting, project implementation, and fundraising.
3. Spread our message and clarify our arguments with public-facing academic deliverables.
4. Build more relationships with the optimal philanthropy, humanist, and critical thinking communities, which share many of our values.
5. Cultivate and expand Singularity Institute's Volunteer Program.
6. Improve Singularity Institute's web presence.
    a. Redesign Singularity Institute's primary website.
    b. Produce more introductory material for the primary website, like the Singularity FAQ.

c. Drive more traffic to Singularity Institute-owned websites.
d. Continue to produce articles on targeted websites and other venues, for example Humanity+, Singularity Hub, IEET, and KurzweilAI.net.
e. Produce high-quality videos to explain Singularity Institute's mission.

7. Apply for grants, especially ones that are given to other organizations and researchers concerned with the safety of future technologies (e.g. synthetic biology and nanotechnology).

8. Continue targeted interactions with the public.
   a. Target media opportunities like NPR's *All Things Considered*, *Gizmodo*, and some documentary crews.
   b. Continue production of documents for wider audiences, for example Yudkowsky's books on rationality.

9. Improve interactions with current and past donors.
   a. Keep our database of all current and past donors updated.
   b. Communicate more regularly with our donor base.
   c. Recognize public donors by maintaining an up-to-date donor list on our website.

Strategy #3: **Improve the function and capabilities of the organization.**

1. Encourage a new organization to begin rationality instruction similar to what Singularity Institute did in 2011 with Rationality Minicamp and Rationality Boot Camp.

2. Use Charity Navigator's guidelines to improve financial and organizational transparency and efficiency.
   a. Use a Transparency or About Us web page to publish Form 990s and other documents showing sound organizational management.
   b. Grow to five independent board members.
   c. Publish audited financial documents prepared by an independent accountant.
   d. Publish a conflict of interest policy.
   e. Publish a whistleblower policy.
   f. Publish a records retention policy.
   g. Publish a policy for determining key officers' compensation.
   h. Publish a privacy policy.
   i. List board members on our website.

3. Ensure a proper orientation for new Singularity Institute staff and visiting fellows.

4. Secure lines of credit to increase liquidity and smooth out the recurring cash-flow pinches that result from having to do things like make payroll and rent event spaces.

5. Improve safe return on financial reserves.

6. Ensure high standards for staff effectiveness.
   a. Provide all staff with a job description and temporally defined goals.
   b. Have employees maintain work logs that are viewable by the board and by executives.

7. When hiring, advertise for applications to find the best candidates.

Near-Term Priorities

Our **top priorities** for 2011-2012 are:

- **Public-facing research on creating a positive singularity**. Publishing more research relevant to creating a positive singularity will make direct progress on our core mission and improve the credibility of our organization, which will in turn improve our ability to raise funds and collaborate with the most talented researchers.

- **Outreach / education / fundraising**. The rationality community has been a critical source of funding and talent, and should be cultivated. The optimal philanthropy, humanism, and critical thinking communities should also be engaged because they are often sympathetic to our goals. Yudkowsky's books on rationality will also reach new audiences that may contain the technical researchers we need. These activities can also diversify our funding sources.

- **Improved organizational effectiveness**. Non-profit best practices can be implemented to improve transparency, public communication, organizational efficiency, and donor interaction.

- **Singularity Summit**. The Singularity Summit spreads our core message, raises funds, and sometimes brings new top-level researchers to work on problems related to AI.

Expansion Plans

Level One (fundraising levels remain steady)

- The Singularity Summit can continue to grow without additional funds.
- Yudkowsky will finish writing his rationality books and return to Friendly AI research.
- Redesign website.
- Grow the Research Associates program. Invite 4-8 additional researchers focused on AI, decision theory, and ancillary fields to join our program
- Grow the Volunteer program.
    - Improve volunteer site and make it public.
    - Develop more organizers who build the program, create projects, and oversee work.
    - Engage volunteers in more high-impact fundraising and outreach activities.

Level Two (fundraising grows by $500,000 and remains steady)

- Level One activities.
- Hire two additional full-time researchers.
- Run a two-week summer program for young people who have qualified for the United States Math Olympiad Training Program, the Computing Olympiad Training program, or have similar strong indications of talent (e.g., top 50 on the Putnam). The program will introduce them to anthropics, decision theory, and other relevant research, and allow them to try their hand at research. This may produce useful work on the topic, and may even discover motivated talent that could join our research team.
- Offer large financial prizes for solving important problems related to our core mission.
- Allow tenured professors to apply to do research with us as visiting professors during their sabbatical.
- Run workshops or seminars on topics related to AI, perhaps in partnership with other sympathetic organizations like the Future of Humanity Institute.