THE SINGULARITY SUMMIT *2010*

August 14–15, 2010
Hyatt Regency • San Francisco, CA

What will happen when and if greater-than-human intelligence is created, either by enhancing the human brain or building an artificial intelligence? Already, technological advances pave the way for further technological advances at an ever-accelerating rate; if intelligence itself could be brought to engineer still greater intelligence in a strong positive feedback loop, the consequences (for better or for worse) would be unimaginable. Author and computer scientist Vernor Vinge called such an event a **Singularity**, alluding to a breakdown in our model of the future, much as how ordinary physical assumptions fail to hold at the center of a black hole.

In its fifth year, the **Singularity Summit** returns to the San Francisco Bay Area to again cast a hopeful yet cautious eye on the latest developments in science and technology and what they portend for the future of intelligence, and thus, the future of humanity.

ABOUT THE SINGULARITY INSTITUTE

The Singularity Institute for Artificial Intelligence is a 501(c)(3) nonprofit dedicated to research and rational deliberation on the future of humanity and in particular the promise and perils of advanced artificial intelligence. We advance this research through our publications, the online community *Less Wrong*, grants, our Visiting Fellows program, and the annual Singularity Summit.

Singularity Institute Research Fellow Eliezer Yudkowsky pioneered the study of AI architectures aimed specifically at ethical, carefully-controlled self-improvement. His papers and essays such as "Staring into the Singularity," "Creating Friendly AI," "Cognitive Biases Potentially Affecting Judgment of Global Risks" and "Artificial Intelligence as a Positive and Negative Factor in Global Risk," introduced a generation to related issues. He is currently writing a book on reductionism, techniques for avoiding common human biases, and other means of thinking productively about AI despite the vast unknowns.

The Institute's Visiting Fellows program just passed its one-year anniversary, and has now brought over thirty graduate students, faculty, and others for short stays in which to learn, trade ideas, and contribute to research and other projects. The Fellows' recent academic publications include "Convergence of Expected Utility for Universal Artificial Intelligence," "Ontological Crises in Artificial Agents' Value Systems," "From Mostly Harmless to Civilization-Threatening: Pathways to Dangerous Artificial Intelligences," and "How Intelligible is Intelligence?" Alumni have gone on to give academic presentations on Coherent Extrapolated Volition, "Economic Implications of Software Minds," "Observer Selection Effects and the Feasibility of Artificial Intelligence," and other relevant research; to start *Less Wrong* meetups in their local communities; and to generally seed careful thinking on long-term AI outcomes.

SCHEDULE

Saturday, August 14

8:00 AM **Breakfast:** coffee and bagels served by Boudin (see map, p.12; volunteers will assist with directions)

9:00 AM Introduction
Sean McCabe

9:05 AM The Darwinian Method
Michael Vassar, Singularity Institute

9:45 AM Evolution and the Posthuman Future
Gregory Stock, Signum Biosciences

10:30 AM **Break:** coffee served at Boudin

11:00 AM The Mind and How to Build One
Ray Kurzweil (live videoconference)

12:05 PM AI Against Aging
Ben Goertzel, Novamente

12:40 PM **Lunch:** at Boudin

2:10 PM Extending Ourselves With Technology
Steve Mann, Univ. of Toronto

2:50 PM Enhancing Our Bodies and Evolving Our Brains
Mandayam A. Srinivasan, MIT

3:25 PM The Past, Present, and Future of Brain-Computer Interfacing
Brian Litt, Univ. of Pennsylvania

4:15 PM Machine Learning Is Rapidly Discovering How the Brain Works
Demis Hassabis, Univ. College London

4:50 PM **Break:** coffee served at Boudin

5:20 PM Reverse-engineering Brains Within Reach
Terrence J. Sejnowski, Salk Institute

5:50 PM What Cells Can Do That Robots Can't
Dennis Bray, Univ. of Cambridge

6:15 PM Debate: Will We Soon Realistically Emulate Biological Systems?
Terrence J. Sejnowski & Dennis Bray

7:00 PM *First day ends*

Sunday, August 15

8:00 AM **Breakfast:** coffee and bagels served by Boudin (see map, p.12; volunteers will assist with directions)

9:00 AM Simplified Humanism and Positive Futurism
Eliezer Yudkowsky, Singularity Institute

9:40 AM The Digital Biome
Ramez Naam

10:15 AM **Break:** coffee served at Boudin

10:40 AM Modifying the Boundary Between Life and Death
Lance B. Becker, Univ. of Pennsylvania

11:15 AM The MRL Mouse: How It Regenerates—and How We Might Do the Same
Ellen Heber-Katz, Wistar Institute

11:50 PM Information Processing in Nanomachines That Read and Write DNA
Anita Goel, Nanobiosym

12:35 PM **Lunch:** at Boudin

2:00 PM Universal Measures of Intelligence
Shane Legg, Univ. College London

2:40 PM Can Discovering the Design Principles of Natural Intelligence Unleash Breakthroughs in AI?
John Tooby, UCSB

3:20 PM Panel Discussion: Narrow and General Intelligence
John Tooby, Ben Goertzel, Eliezer Yudkowsky, and Shane Legg

4:00 PM **Break:** coffee served at Boudin

4:25 PM Emotionally Intelligent Machines
David Hanson, Hanson Robotics

5:00 PM The Future of Energy
José Luis Cordeiro

5:35 PM Nonhuman Intelligence: Where We Are and Where We're Headed
Irene Pepperberg, Alex Foundation

6:15 PM Is There Such a Thing as Scientific Consensus?
James Randi, JREF

7:00 PM *Summit ends*

PRESENTATION ABSTRACTS

The Darwinian Method
Michael Vassar

People have always acquired useful knowledge about the world around them, but there is an important sense in which science only began a few hundred years ago. We resolve this apparent paradox by recognizing that humans have developed many different tools to allow reason to guide their collective actions. One such tool, what I call "Enlightenment Science," grew out of philosophical skepticism to transform the world quite recently. Many scientific controversies can best be understood as conflicts between those who accept an earlier scientific method, that of "Scholarly Science," and those who believe "Enlightenment Science" to be the only source of authoritative knowledge. In truth, both scholarly and enlightenment scientific methods have serious weaknesses, and much important science depends on a proper synthesis of the two methods, as does the argument for anticipating a technological Singularity. I will explain why these two scientific methods work, where they come into conflict, and how such conflicts can be resolved.

Evolution and the Posthuman Future
Gregory Stock

We are in the midst of a major evolutionary breakthrough. Life is beginning not only to reshape itself and its workings, but to animate the inanimate world around it. The idea that human values and consciousness might transition to post-singularity cyberspace and survive in a realm populated by super-AIs and brain emulations is alluring to many. But a macro-evolutionary perspective suggests that any uploaded "human" intelligence would be far more likely to mutate into something "other." Our nature and values evolved through selective pressures that enhanced reproductive success among primates. Human values would make little sense for AIs living where backups and copies are easy, personal boundaries are weak, and sexual reproduction is absent. Evolution will drive such cyber beings towards values more suited to their circumstances and environment.

The Mind and How to Build One
Ray Kurzweil

As we learn the algorithms underlying human intelligence, we will similarly be able to engineer it to vastly extend the powers of our intelligence. Indeed this process is already well under way. There are literally hundreds of tasks and activities that used to be the sole province of human intelligence that can now be conducted by computers usually with greater precision and vastly greater scale. According to my models, we are only two decades from fully modeling and simulating the human brain. By the time we finish this reverse-engineering project, we will have computers that are millions of times more powerful than the human brain. These computers will be further amplified by being networked into a vast world wide cloud of computing. The algorithms of intelligence will begin to self-iterate towards ever smarter algorithms. It is only by extending our intelligence with our intelligent technology that we can handle the scale of complexity to address the challenges our civilization faces.

AI Against Aging
Ben Goertzel

Death due to aging is a solvable problem. The human body is a complex machine, modifiable and reparable like any other machine. The worst bottleneck we face in preventing aging is the effectiveness of human researchers: the human brain simply was not evolved for the integrative analysis of a massive number of complexly-interrelated, high-dimensional biological datasets. The most feasible solution, then, is to supplement human biological scientists with increasingly advanced artificial intelligence. The Biomind AI tools my colleagues and I have developed have helped discover the genetic basis of chronic fatigue syndrome, and to create novel diagnostics for Parkinson's disease based on identifying subtle patterns of damage in mitochondrial DNA. Today's AI systems are already comprehending the biological world in ways far beyond the human brain's capability, but they illustrate only a tiny fraction of what's to come.

Enhancing Our Bodies and Evolving Our Brains
Mandayam A. Srinivasan

Although machines vary, they ultimately serve one purpose: to extend our ability to interact with the world around us. Rapid evolution of technology is bringing humans and machines ever closer, even occasionally merging the two within our bodies to augment our biological functions. Interactions with real and virtual

environments through sophisticated machines can enhance both our sensorimotor and cognitive abilities, thus contributing not only to the extension of our bodies but also to the evolution of our brains.

Human-machine interaction includes two-way transmission of information: sensory information from the environment to the human that affects our perception of the environment, and action commands from the human to the environment to explore or modify the environment. Current technologies that interface with our haptic system show the path towards removing some of the limitations of our bodies in physically acting on the environment. By providing unprecedented stimuli and responses in real and virtual worlds, these machines enable evolution of our brains as well.

Brain-Computer Interfaces: Past, Present, and Future
Brian Litt

Brain-Computer Interfaces (BCI) interact with neuronal networks to restore function and treat human disease. Devices for Parkinson's disease, pain, epilepsy and deafness are now becoming mainstream, with hundreds of thousands of people implanted each year. Other applications such as blindness, paralysis, depression and memory loss aren't far behind, with unprecedented opportunity to improve our quality of life. Currently BCIs are limited by the number, scale and resolution of brain sensors, our ability to map brain circuits, and our capacity to interpret and respond to Terabyte-scale data streams in real time. Converging advances in engineering, medicine and neuroscience are addressing these problems and accelerating BCI research. It is only a matter of time before BCI development moves from circuit repair to augmenting human function. It is imperative that we think carefully and scientifically about how this shift should take place, and that we be prepared to guide this process.

Combining Systems Neuroscience and Machine Learning: a New Approach to AGI
Demis Hassabis

Neuroscience is rapidly teasing apart the functional roles of the brain's components, and in some cases even the types of algorithms that they use. Machine learning, meanwhile, is producing a growing collection of techniques for specific kinds of problems, but as yet no general purpose algorithm for artificial intelligence. By bringing these two fields together, we can have both a high level architecture for an artificial general intelligence, and working algorithms for implementing many of the required components. I will outline the case for pursuing this approach, some current work in progress, and some of the challenges we face going forward.

Reverse-engineering the Brain Is Within Reach
Terrence J. Sejnowski

When we come to understand how brains function, we should become able to build devices with cognitive abilities—such as cognitive cars that are better at driving than we are because they communicate with other cars and share knowledge on road conditions. Early versions of intelligent communications systems are already planned for the next federal auction of the electromagnetic spectrum, and soon to come are similar ways to enhance other utilities, such as the "cognitive power grid," which will automatically anticipate and regulate the flow of power around the country. Sensors will stream information—on the use of electricity, weather patterns, and travel conditions—and use this information to optimize goals, such as reducing power usage and travel time, by regulating the flow of resources. Parts of this system are already in place, such as sensors and the internet, but there is as yet no central nervous system to integrate this torrent of information and take appropriate actions. As it increasingly mimics the workings of our brains, the world around us will become smarter and more efficient.

What Cells Can Do That Robots Can't
Dennis Bray

Though they are superb specialists able to do one thing supremely well, robots and their computer brains lack the adaptability and the self-regenerating generalist abilities of living organisms. So what is missing? The answer, I argue, can be found in the discoveries of contemporary biology. Living cells are crammed full of molecules that act as biochemical switches. Most are allosteric and modifiable—formally equivalent to transistors—and linked into extensive networks through diffusion-limited binding events and biochemical reactions. But although biological components act in ways that are comparable to those in electronic circuits, they are set apart by the huge number of different states they can adopt. Multiple biochemical processes create chemical modifications of protein molecules, further diversified by association with distinct structures at

defined locations of a cell. The resulting combinatorial explosion of states endows living systems with an almost infinite capacity to store information regarding past and present conditions and a unique capacity to prepare for future events.

Simplified Humanism and Positive Futurism
Eliezer Yudkowsky

I will discuss a complex of related ideas that together form a more powerful whole: The idea of an upward-sloping future, as opposed to a worldview that locates utopia in the past and sees an inevitable downward slope going forward; humanism that focuses on personhood theory, rather than matters of outward form; the rejection of previously accepted ills, such as smallpox, or now aging, as unjustifiable; commitments to reason and rationality, the substitution of goals for prophecies, and the acceptance that each additional detail of a prediction is burdensome and must be separately justified. These ideas combine to form the natural continuation of the Enlightenment project, in which hope for a dramatically better future continues.

The Digital Biome
Ramez Naam

Exponential technologies offer the promise of radically altering the face of the planet on which we dwell. Within the next 20 years we will have sequenced the genome of every known species on the earth and tremendously advanced our understanding of how to utilize those genes and reprogram those organisms to alter the biosphere. Biosphere engineering will play a major role in overcoming current environmental and resource challenges, including finite reserves of fossil fuels and looming changes to the earth's climate. With effective technology to sculpt the planetary biome, the limits of the number of humans that can live on the planet, and the quality of life of each, will be tremendously higher than they appear to be today. This talk will explore some of the lower bounds of what's possible with control of the biome.

Modifying the Boundary Between Life and Death
Lance B. Becker

There is increasing evidence that the boundary between life and death is not a distinct cut-off, but rather is more flexible and modifiable than we have previously believed. If true, it may be possible to save tens of thousands of lives each year from premature death with novel therapies currently in development.

A common pathway for sudden death is any condition that results in *ischemia*, or oxygen deprivation to tissues. Rapid *reperfusion*—the restoration of oxygenated blood—is currently the universal treatment for life-threatening emergencies involving ischemia, but this approach causes tissue injury and does not work if the period of oxygen deprivation has been prolonged.

In recent years, we have developed some newer insights into the basic mechanisms of cell death and the cause of reperfusion injury. There is reason for much optimism that we will soon have the ability to restore life to tens of thousands of premature deaths each year with these new approaches to resuscitation.

The MRL Mouse: How It Regenerates and How We Might Do the Same
Ellen Heber-Katz

I will present an overview of the current state of mammalian regeneration in the context of recent and ongoing discoveries in the MRL mouse system. The MRL mouse is the first and to date the only genetically dissectable mammalian model for regeneration and as such provides a "laboratory" for assessing theories and potential clinical approaches to achieving the medical goal of regeneration. The very recent finding that the turning off of a single gene—*p21*—related to the regulation of the cell cycle can convert an otherwise non-regenerating mouse to a regenerator leads to two important conclusions. First, the molecular apparatus for regeneration has survived the long evolutionary gap between lower species such as hydra and salamander up through the advent of mammals. And second, the fact that the capacity to regenerate can be "uncloaked" through the turning off of a single gene gives impetus to the possibility that regeneration can be induced by drug therapy alone.

Universal Measures of Intelligence
Shane Legg

Relatively little work has gone into developing general, encompassing and theoretically founded definitions of intelligence for machines. This leaves us without a clear foundation for either theoretical research or developing empirical measures of progress. I outline the major perspectives on the nature of intelligence and some of the informal definitions that have been put forward. I then sketch the main ideas behind the universal intelligence measure, a formal definition of

intelligence based on Hutter's AIXI model of theoretically optimal machine intelligence. Based on this a number of researchers, including myself, are developing practical tests of machine intelligence. I describe some of the challenges faced when doing this and share some recent results from testing various artificial agents.

Can Discovering the Design Principles Governing Natural Intelligence Unleash Breakthroughs in Artificial Intelligence?

John Tooby

The success of evolutionary psychology will be pivotal for the program of building a useful broad artificial intelligence. To begin with, for any such nonhuman intelligence to communicate with humans or respect human values, it will have to become equipped with accurate models of the representational and motivational programs that inhabit human minds. More importantly, human intelligence offers us a working prototype of computational intelligence: reverse-engineering the human computational architecture and uncovering its design principles provides key insights into how natural selection achieved computational broad intelligence.

It has traditionally seemed self-evident to AI researchers that general intelligence must be achieved by general-purpose algorithms, yet the recurrent design principle emerging from evolutionary psychology is that the natural intelligences found in humans operate, with few exceptions, by being specialized. Natural selection breaks off small but biologically important fragments of the universe and engineers distinct problem-solving methods for each. This suggests that an alternative and more fruitful road to engineering a broad artificial intelligence is not by searching for algorithms that manifest general intelligence, but instead by aggregating specialized intelligences.

Emotionally Intelligent Machines

David Hanson

The human mind is hardwired for social cognition and face-to-face interactions. This explains people's powerful attraction to fictional characters in entertainment and literature. The effect could extend to AI-driven character machines, if we can make their intelligence humanlike enough to act like protagonists. Currently robots and agents may look extremely humanlike, but possess only rudimentary AI. To realize

the potential for enchanting, lovable protagonists, AI must closely emulate human emotions and cognition. The machines must truly understand us and exhibit genuine compassion and creativity. As protagonist machines obtain these capabilities, they will grow more desirable to consumers. This feedback loop among consumer demand for lovable characters, and technical development of products that better satisfy people, could propel the evolution of strong, friendly AI. Hence, humanlike character intelligence holds promise as an approach for the pursuit of safe, singularity-level machine intelligence.

Nonhuman Intelligence: Where We Are and Where We're Headed

Irene Pepperberg

Futurists like to predict how genetic engineering and computational implants will allow humans to become a super-species, but what about the application of similar technologies to nonhumans? First, however, shouldn't we examine their current cognitive and physical capacities? How can we improve upon what we don't yet fully understand? Second, what would we do with more intelligent creatures—harness them to do more of our work or treat them more as equals? Finally, we already have many conflicting views on the legal status of animals in our society—what happens if we endow them with more human characteristics? Might such animals provide a testing ground for humanity's ability to interact humanely with animal-like and ultimately human-like machines?

Is There Such a Thing as Scientific Consensus?

James Randi

We like to believe we can trust our perceptions, but in my travels as a conjurer, escape artist, and investigator of paranormal claims, I have learned much about how human beings fool each other and themselves. The inhabitants of the modern world are not as rational as they appear, and as our technologies become ever-more potent, our hidden penchant for unreason becomes commensurately more dangerous. In singular times, it is the ethical responsibility of every thinking being to become an agent for the promulgation of critical thought, skepticism, and humility.

SPEAKER BIOGRAPHIES

Lance B. Becker is Professor of Emergency Medicine and Director of the Center for Resuscitation Science at the University of Pennsylvannia medical center. Becker has published dozens of papers in peer-reviewed journalsm is a member of the Institute of Medicine (one of the highest honors bestowed in the biomedical field), and has received numerous awards. Becker's research has helped define cellular reperfusion injury mechanisms, mitochondrial medicine, and hypothermia protection. Ongoing studies include development of novel human coolants for rapid induction of hypothermia, improving the quality of CPR, new defibrillator and cardiopulmonary bypass technologies, and epidemiology of sudden death.

Dennis Bray is an active emertius professor at the University of Cambridge. Previously accomplished in neurobiology, he now works on computational models of cell signaling. He is the author of the 2009 book *Wetware: A Computer in Every Cell*. He coauthored the 2003 textbook *Essential Cell Biology* as well as previous editions of the classic textbook *Molecular Biology of the Cell*.
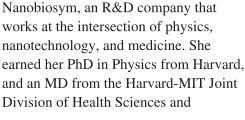
José Luis Cordeiro is director of the Venezuela node of the Millenium Project. He worked as an engineer in petroleum exploration for the French company Schlumberger. For over six years, he served as an advisor for many of the major oil companies in the world, including Agip, BP, ChevronTexaco, ExxonMobil, PDVSA, Pemex, Repsol, Shell and Total. Later, in Paris, he initiated his relation with the international consulting company Booz-Allen & Hamilton, where he specialized in the areas of strategy, finance and restructuring.

Anita Goel is founder and CEO of Nanobiosym, an R&D company that works at the intersection of physics, nanotechnology, and medicine. She earned her PhD in Physics from Harvard, and an MD from the Harvard-MIT Joint Division of Health Sciences and Technology.

Ben Goertzel is CEO and Chief Scientist of AI firm Novamente, LLC, as well as CEO of bioinformatics firm Biomind LLC. A research faculty member for eight years in several universities in the US and Australasia, he has published over 80 research papers in in disciplines spanning AI, mathematics, computer science, cognitive science, philosophy of mind, and bioinformatics.

David Hanson is a roboticist and founder of Hanson Robotics. He works on designing and developing robots with human-like expressive capabilities. Hanson holds a patent on *Frubber*, a novel material that imitates the look and feel of human skin, used in the creation of androids.

Demis Hassabis is a reasearch fellow at the Gatsby Computational Neuroscience Unit at University College London. There, he investigates the mechanisms underlying episodic memory, imagination, and planning. Prior to his work in neuroscience, Hassabis was a successful video game designer and artificial intelligence programmer, having founded and run the game development firm Elixir Studios.

Ellen Heber-Katz is a professor at the Wistar Institute in Philadelphia. Her research focuses on the molecular biology and genetics of healing and the autoimmune system, basic research that is being applied to meningitis, heart disease, spinal cord damage, and other human maladies. Heber-Katz earned her MS from the University of Wisconsin in 1972 and her PhD from the University of Pennsylvania in 1976.

Ray Kurzweil is one of the leading inventors of our time. He was the principal developer of the first CCD flat-bed scanner, the first omni-font optical character recognition, the first print-to-speech reading machine for the blind, the first text-to-speech synthesizer, the first music synthesizer capable of recreating the grand piano and other orchestral instruments, and the first commercially marketed large-vocabulary speech recognition device. Kurzweil's website *KurzweilAI.net* has over one million readers. Kurzweil has written six books, four of which have been national bestsellers.

Shane Legg did his PhD work on theoretical models of superintelligent machines with Prof. Marcus Hutter at the Swiss IDSIA research lab. This research was hailed in international conferences and journals, and upon the completion of his PhD, Legg was awarded the 2008 Canadian Singularity Institute for Artificial Intelligence Research Prize worth $10,000. He is currently working with Prof. Peter Dayan at the Gatsby Computational Neuroscience Unit, University College London. Previously, he has done research at the University of Waikato, the University of Aukland, and worked as an AI programmer and software engineer for Webmind and Adaptive Intelligence.

Brian Litt is a professor of neurology and bioenginering at the University of Pennsylvania. His research specializes in the care and treatment of individuals with epilepsy.

Steve Mann is a professor at the University of Toronto. He is a pioneer in the study and practice of virtual reality, to the extent that more than one publication has dubbed him the world's first cyborg. He is the author of a popular book on cyborg living, *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Comptuer*. Mann is the author of more than 200 research publications.
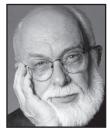
Ramez Naam is a technologist who has worked for Microsoft (working on such popular products as Internet Explorer and Outlook) and served as the CEO of Apex Nanotechnologies, a company that develops software tools for nanotechnological design. He is the author of *More than Human: Embracing the Promise of Biological Enhancement*.

Irene Pepperberg is a professor of psychology at Brandeis University and a world-renowned expert on animal cognition. She is famous for her work with Alex, an African Grey Parrot whom Pepperberg studied and taught to use and understand over 150 words.

James Randi is a retired professional magician ("The Amazing Randi"), skeptic, author, lecturer, and amateur archaeologist and astronomer. He has written a number of popular books on skepticism and stage magic. He was born in 1928 in Toronto, Canada, where he received his high school education. He was naturalized a U.S. citizen in 1987, and now lives in Florida.

Terrence J. Sejnowski is a professor of neurobiology at the University of California at San Diego, the director of the Computational Neurobiology Lab at the Salk Institute for Biological Sciences, and an investigator at the Howard Hughes Medical Institute. His research uses both experimental and modeling techniques to unravel the connections between brain and behavior.

Mandayam A. Srinivasan is a professor in the Department of Mechanical Engineering at the Masseschussetts Institute of Technology. His research focuses on haptics, tje study of tactile interaction between humans and machines. He earned his PhD from Yale University in 1984.

Gregory Stock is a biophysicist, author, and entrepreneur, and presently the CEO of Signum Biosciences, a medical biotechnology company that is developing candidate treatments for Parkinson's disease, Alzheimer's disease, and diabetes. His 2002 book, *Redesigning Humans* is a transhumanist classic, following up on his 1992 work on human-machine symbiosis, *Metaman*. Stock serves on the editorial board of the *International Journal of Bioethics*, as well as three other journals.

John Tooby is one of the pioneers of evolutionary psychology. Tooby has coauthored papers on the foundations of evolutionary psychology and psychological adaptations faciliating social exchange, and coedited the 1992 seminal anthology *The Adapted Mind*. Along with his wife and frequent collaborator Leda Cosmides, he cofounded the Center for Evolutionary Psychology at the University of California at Santa Barbara. He is currently working on a book about the evolution of sexual reproduction as an adaptation to combat parasites.

Michael Vassar is the president of the Singularity Institute, a nonprofit devoted to research about existential risks and recursively self-improving artificial general intelligence, and host of the Singularity Summit. Previously, he was a founder and Chief Strategist at SirGroovy.com, an online music licensing firm, and held positions with Aon, the Peace Corps, and the National Institute of Standards and Technology. Michael has been writing and speaking on topics related to the safe development of disruptive technologies for a number of years. He holds an MBA and a BS in biochemistry.

Eliezer Yudkowsky is an artificial intelligence theorist, prolific writer about human rationality, and cofounder of the Singularity Institute. Much of his writings on rationality are hosted at the popular community blog *Less Wrong*, which has received over 2 million visits. He is currently working on a popular-level book on rationality based on his writings at *Less Wrong*.

SPECIAL THANKS

SPONSOR

Jaan Tallinn &

AMBIENT SOUND INVESTMENTS
www.asi.ee

Ambient Sound Investments (ASI) was established in 2003 as an equal partnership by four founding engineers at Skype Technologies to hold a minority stake in Skype. At the end of 2005 ASI sold its stake to eBay and now operates as a private investment vehicle. Today ASI is a team of about 10 people managing €100 million of the partners' assets and growing an independent investment vehicle spanning multiple generations.
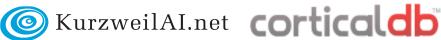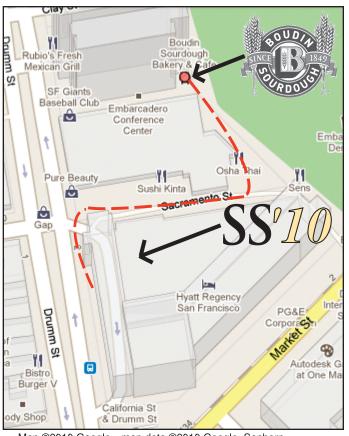
PARTNERS

Also

The Singularity Institute would like to thank everyone else who helped make this Summit possible, including volunteers Frank Adamek, Jolly Apneet, Kenan Azam, Peter de Blanc, Mike Blume, Ariel Berwaldt, Yonah Berwaldt, Tim Czech, Zack M. Davis, Jonathan El-Bizri, Dennis Fan, Emil Gilliam, Andrew J. Hacker, Nick Hay, Louie Helm, Todd Huffman, Jesse Liptrap, Stanislav Luban, Mark Maccabee, Sean McCabe, Thomas McCabe, Alex Pantek, Kala Perkins, Andrew Rettek, Jessica Scorpio, Anthony Scratchell, John Sleeman, and Nick Tarleton.

Map ©2010 Google—map data ©2010 Google, Sanborn

During the breaks and lunch, food will be provided by the ***Boudin Bakery*** outside the Hyatt. Volunteers and signs will help point the way; be sure to wear your Summit name badge.

***Thank you for attending!—
hope to see you next year!***
If you'd like to see the work of the Singularity Institute continue, please consider making a donation:
www.singinst.org/donate

RESOURCES & FURTHER READING

"The Coming Technological Singularity" by Vernor Vinge
In this classic paper that popularized the term *Singularity*, Vinge (writing in 1993) argues that within decades, superintelligence will be developed, thereby ending the human era.
http://www.accelerating.org/articles/comingtechsingularity.html

"Reducing Long-Term Catastrophic Risks From Artificial Intelligence" by the Singularity Institute
This expository article articulates the current state of the artificial intelligence risks landscape and explains its importance.
http://www.singinst.org/riskintro/index.html

"The Hanson–Yudkowsky AI Foom Debate" at the *Less Wrong* wiki
In a series of blog posts at *Less Wrong* and Robin Hanson's

Overcoming Bias blog, Hanson and Singularity Institute research fellow Eliezer Yudkowsky debate the character of artificial intelligence development.
http://wiki.lesswrong.com/wiki/The_Hanson-Yudkowsky_AI-Foom_Debate

"The Basic AI Drives" by Steve Omohundro
Omohundro argues that several basic motives will arise in most sufficiently advanced AIs. AIs with a wide range of terminal values will all want to self-improve, to be rational, and protect themselves.
http://selfawaresystems.files.wordpress.com/2008/01/ai_drives_final.pdf

"Existential Risks: Analyzing Human Extinction Scenarios" by Nick Bostrom
Bostrom classifies and analyzes major threats to the human species and its potential.
http://www.nickbostrom.com/existential/risks.html