One Hundred Year Study on Artificial Intelligence

Barbara J. Grosz, Chair AI100 Standing Committee
Harvard University

Envisioning a long horizon

An enduring institution

Create an Enduring Process

Motivation: It is difficult to anticipate the opportunities AI offers to make a difference to society or the issues realizing those ambitions will raise.

"It's our hope that the study, with its extended memory and long gaze, will provide important insights and guidance over the next century and beyond."

Eric Horvitz

One Hundred Year Study
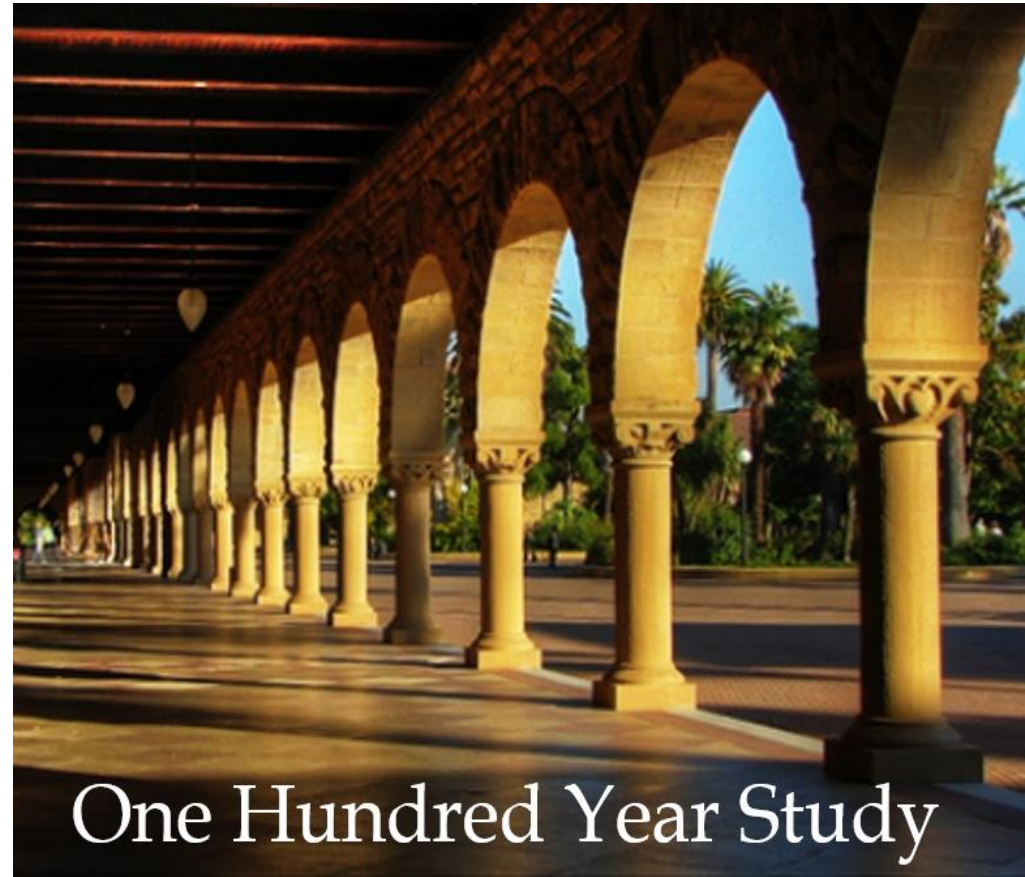
Roots: AAAI Presidential Panel on Long-Term AI Futures (2008-09)

aka "the AAAI Asilomar Study"

Commissioned by AAAI President, Eric Horvitz

Co-chaired by Eric Horvitz & Bart Selman

http://www.aaai.org/Organization/presidential-panel.php

Presidential Panel on Long-Term AI Futures

Charge: Explore the potential long-term societal influences of AI advances.

\

Subgroups:

- Potential Disruptive Advances Over the Short-term

- Longer-term Pace, Concerns, Control

- Ethical and Legal Challenges

Presidential Panel on Long-Term AI Futures

Charge: Explore the potential long-term societal influences of AI advances.

\

Subgroups:

- Potential Disruptive Advances Over the Short-term (Milind Tambe)

- Longer-term Pace, Concerns, Control

- Ethical and Legal Challenges

Presidential Panel on Long-Term AI Futures

Charge: Explore the potential long-term societal influences of AI advances.

Subgroups:

- Potential Disruptive Advances Over the Short-term

- Longer-term Pace, Concerns, Control (Dave McAllester)

- Ethical and Legal Challenges

Presidential Panel on Long-Term AI Futures

Charge: Explore the potential long-term societal influences of AI advances.

\

Subgroups:

- Potential Disruptive Advances Over the Short-term

- Longer-term Pace, Concerns, Control

- Ethical and Legal Challenges (Dave Waltz)

Findings: Value of Endeavor

- Valuable endeavor: assessing the field and its potential
  - Trends
  - Potential disruptions
  - Societal influences and ethical issues

- Many opportunities, many challenges

- Need to engage within AI community and to the outside

Findings: Value of Endeavor

- Valuable endeavor: assessing the field and its potential
  - Trends
  - Potential disruptions
  - Societal influences and ethical issues

- Many opportunities, many challenges

- Need to engage within AI community and to the outside

Important for AI scientists to be proactive and provide guidance.

Cross disciplinary expertise required to address these issues.

Findings: Value of Endeavor

- Valuable endeavor: assessing the field and its potential
  - Trends
  - Potential disruptions
  - Societal influences and ethical issues

- Many opportunities, many challenges

- Need to engage within AI community and to the outside

Important for AI scientists to be proactive and provide guidance.

Cross disciplinary expertise required to address these issues.

Sense of clear value in repeating this exercise.

Findings: Opportunities & Challenges

- Technical surprises that might lead to jump in capabilities (e.g., "deep learning").
- Potential need & benefits to more focus on collaboration of human and machine.
- Need for better understanding of specification, robustness, control of system behavior.
- Challenges
  - "criminal AI"
  - disruptions

Findings: Engaging inward and outward

Field must pay attention to rising concerns about risks of superintelligences and robots in society.

To address, look
Inward: Differences of opinion within AI community about potential for disruption, loss of control, dystopian/utopian futures.

Outward: Communication with media and to the public directly requires thought and care.

AAAI Presidential Panel on Long-Term AI Futures: An Outside View

The New York Times

Science

Scientists Worry Machines May Outsmart Man

By JOHN MARKOFF
Published: July 25, 2009

A robot that can open doors and find electrical outlets to recharge itself. Computer viruses that no one can stop. Predator drones, which, though still controlled remotely by humans, come close to a machine that can kill autonomously.

TWITTER

LINKEDIN

SIGN IN TO E-MAIL

PRINT

Asilomar study report: http://www.aaai.org/Organization/presidential-panel.php

AAAI Presidential Panel on Long-Term AI Futures: An Outside View

The New York Times

Science

Scientists Worry Machines May Outsmart Man

By JOHN MARKOFF

The A.A.A.I. report will try to assess the possibility of "the loss of human control of computer-based intelligences." It will also grapple, Dr. Horvitz said, with socioeconomic, legal and ethical issues, as well as probable changes in human-computer relationships. How would it be, for example, to relate to a machine that is as intelligent as your spouse?

Dr. Horvitz said the panel was looking for ways to guide research so that technology improved society rather than moved it toward a technological catastrophe. Some

Asilomar study report: http://www.aaai.org/Organization/presidential-panel.php

AAAI Presidential Panel on Long-Term AI Futures: An Outside View

The New York Times   **Science**

Scientists Worry Machines May Outsmart Man

By JOHN MARKOFF

The meeting on artificial intelligence could be pivotal to the future of the field. Paul Berg, who was the organizer of the 1975 Asilomar meeting and received a [Nobel Prize](#) for chemistry in 1980, said it was important for scientific communities to engage the public before alarm and opposition becomes unshakable.

"If you wait too long and the sides become entrenched like with G.M.O.," he said, referring to genetically modified foods, "then it is very difficult. It's too complex, and people talk right past each other."

Asilomar study report: http://www.aaai.org/Organization/presidential-panel.php

One Hundred Year Study

"Artificial intelligence is one of the most profound undertakings in science, and one that will affect every aspect of human life.

"Given's Stanford's pioneering role in AI and our interdisciplinary mindset, we feel obliged and qualified to host a conversation about how artificial intelligence will affect our children and our children's children."

Stanford President John Hennessy
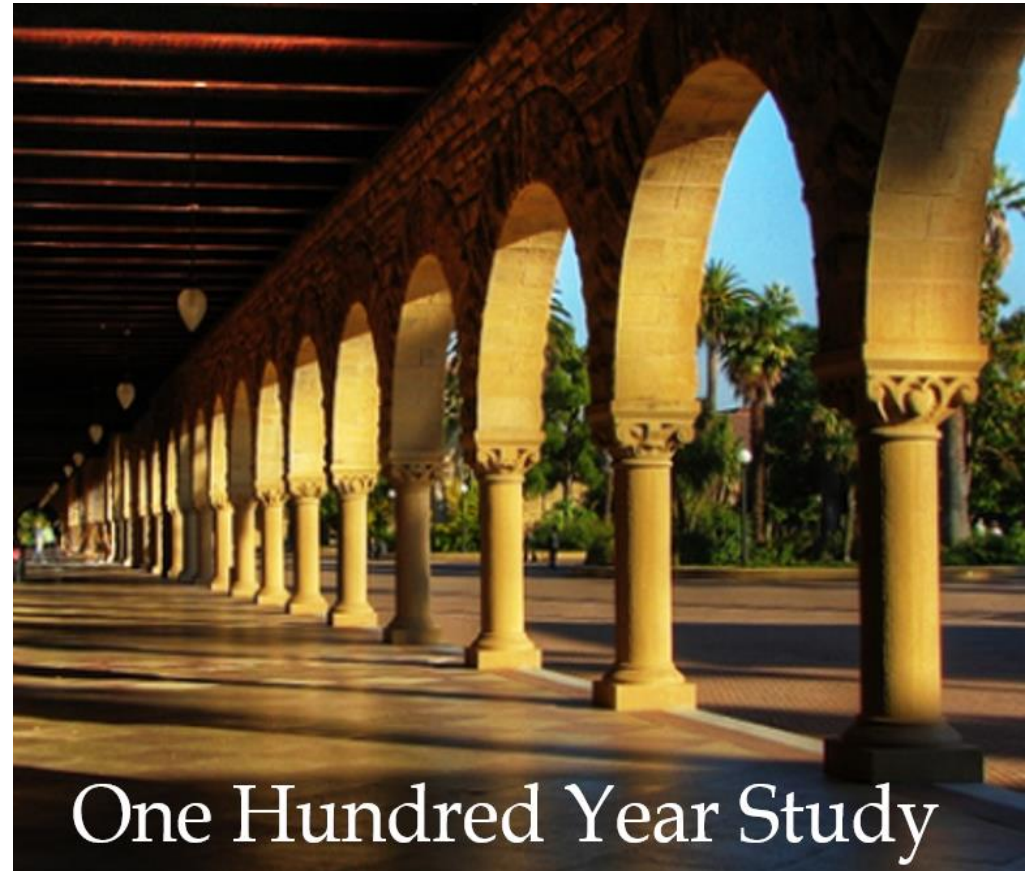
One Hundred Year Study

Launch of the Project: Donors' "Framing Memo" Focal Topics

Technical trends and surprises

Key opportunities for AI

Delays with moving AI into world

Privacy and machine intelligence

Democracy & freedom

AI advances & Law

AI advances & Ethics

AI & Economics

AI & warfare

Collaborations with machines

AI and human cognition

Criminal uses of AI

Safety & autonomy

Loss of control of AI systems

Psychology of people & smart machines

Communication, understanding, outreach

Neuroscience and AI

AI and philosophy of mind

Launch of the Project: Donors' "Framing Memo" Focal Topics

Technical trends and surprises

Key opportunities for AI

Delays with moving AI into world

Privacy and machine intelligence

Democracy & freedom

AI advances & Law

AI advances & Ethics

AI & Economics

AI & warfare

Collaborations with machines

AI and human cognition

Criminal uses of AI

Safety & autonomy

Loss of control of AI systems

Psychology of people & smart machines

Communication, understanding, outreach

Neuroscience and AI

AI and philosophy of mind

Key opportunities for AI: Where Might AI Make the Greatest (Positive) Difference?

- How might AI advances and systems based on them help transform education, healthcare transportation, science, or government, thus contributing to the overall vitality of society?
- What problems and bottlenecks are ripe for being solved or addressed with computational systems that can perceive, learn, reason, and plan?
- Where might AI systems unlock great value by disrupting the status quo in significant ways?

Key opportunities for AI: Where Might AI Make the Greatest (Positive) Difference?

- How might AI advances and systems based on them help transform education, healthcare transportation, science, or government, thus contributing to the overall vitality of society?

- What problems and bottlenecks are ripe for being solved or addressed with computational systems that can perceive, learn, reason, and plan?

- Where might AI systems unlock great value by disrupting the status quo in significant ways?

What major positive changes in the world can AI advances enable?

Meeting the Challenge of Moving AI Into the World

- AI has the potential to improve performance, reduce costs, introduce new efficiencies, and raise the quality of life.
- Already, in many cases, approaches shown to work ***have not come into wide use.***
    - health care: avoidable errors, predicting infection (and taking action)
    - transportation: alerting distracted drivers
- Address the urgent need to better understand how we can more quickly translate valuable existing AI competencies and advances into real-world practice.

Meeting the Challenge of Moving AI Into the World

- AI has the potential to improve performance, reduce costs, introduce new efficiencies, and raise the quality of life.
- Already, in many cases, approaches shown to work **have not come into wide use.**
  - health care: avoidable errors, predicting infection (and taking action)
  - transportation: alerting distracted drivers
- Address the urgent need to better understand how we can more quickly translate valuable existing AI competencies and advances into real-world practice.

How does sluggish translation of AI technologies into the world translate into unnecessary deaths and costs? What can be done to overcome this problem?

Aligning Systems Resulting From AI Advances With Ethics

- Identify ethical challenges and questions that might arise from advances in competencies and uses of AI systems that make inferences and take actions in the world.

    - What uses of AI might be considered unethical?

    - What ethical questions arise from the use of autonomous decision systems in such high-stakes areas as healthcare and transportation?

    - What ethical issues arise from building systems that display human-like qualities and competencies?

Aligning Systems Resulting From AI Advances With Ethics

- Identify ethical challenges and questions that might arise from advances in competencies and uses of AI systems that make inferences and take actions in the world.

  AI advances likely to raise new ethical questions.

  - What uses of AI might be considered unethical?

  - What ethical questions arise from the use of autonomous decision systems in such high-stakes areas as healthcare and transportation?

  - What ethical issues arise from building systems that display human-like qualities and competencies?

Meeting the Challenge of Effectively Combining Human and Computer Expertise and Skill

- Mixed human-computer collaborations and mixed-initiative systems rely on reasoning, guidance, action of both people and AI systems, raising many challenges:
  - identify appropriate ways of mixing and joining efforts
  - address challenges of transfer of control
  - endow machines with the skills to explain their reasoning so that people can understand their inferential steps, conclusions, and recommendations.

Meeting the Challenge of Effectively Combining Human and Computer Expertise and Skill

- Mixed human-computer collaborations and mixed-initiative systems rely on reasoning, guidance, action of both people and AI systems, raising many challenges:

  Fundamentally much harder than coordinating purely computer system teams.

  - identify appropriate ways of mixing and joining efforts
  - address challenges of transfer of control
  - endow machines with the skills to explain their reasoning so that people can understand their inferential steps, conclusions, and recommendations.

Meeting the Challenge of Effectively Combining Human and Computer Expertise and Skill

- Mixed human-computer collaborations and mixed-initiative systems rely on reasoning, guidance, action of both people and AI systems, raising many challenges:

  Fundamentally much harder than coordinating purely computer system teams.

  - identify appropriate ways of mixing and joining efforts
  - address challenges of transfer of control
  - endow machines with the skills to explain their reasoning so that people can understand their inferential steps, conclusions, and recommendations.

Requires cross-disciplinary collaborations and sophisticated empirical work as well as technical advances.

Addressing the Challenge of Autonomy: Safety and Guaranteeing Behavior

- How might we
  - ensure appropriate system behavior, even when autonomous systems encounter new or unforeseen situations in the open world?
  - enable systems to understand when they do not know enough to act in the world and effectively transfer control or seek human advice?
  - enable specification and verification of ranges of desired or safe behaviors of autonomous systems?

Specifying and ensuring the behavior of autonomous systems, especially those that interact with people, is fundamentally hard, much harder than doing so for closed systems.

Communication, understanding, and outreach.

- Addressing the multiple *risks of poor understanding*?
  - How might anxiety about the loss of control of machines slow down beneficial applications of AI methods?
  - How might it lead to a shutdown of important AI research?
- What education and communication programs would be valuable for *informing non-experts* about AI research and capabilities—and the challenges and opportunities ahead?

Many non-expert audiences (media, politicians, scientists in other fields, the "lay public"), each with different understandings and perspectives on the world which they bring to their (mis)understanding of AI.

AI scientists and developers of new AI-based technologies must bridge the gap.

AI 100 Standing Committee

Barbara Grosz

Alan Mackworth

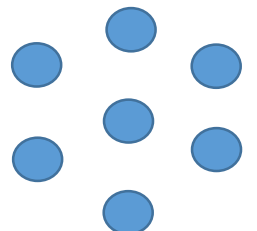Tom Mitchell

Deirdre Mulligan

Yoav Shoham

Eric Horvitz
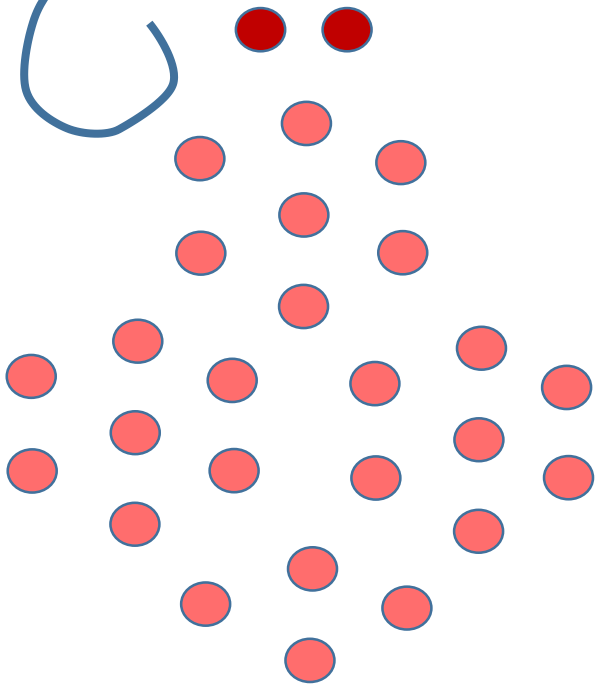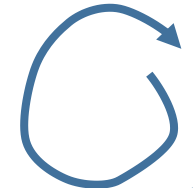

Russ Altman

AI 100 Syncopated Timeline
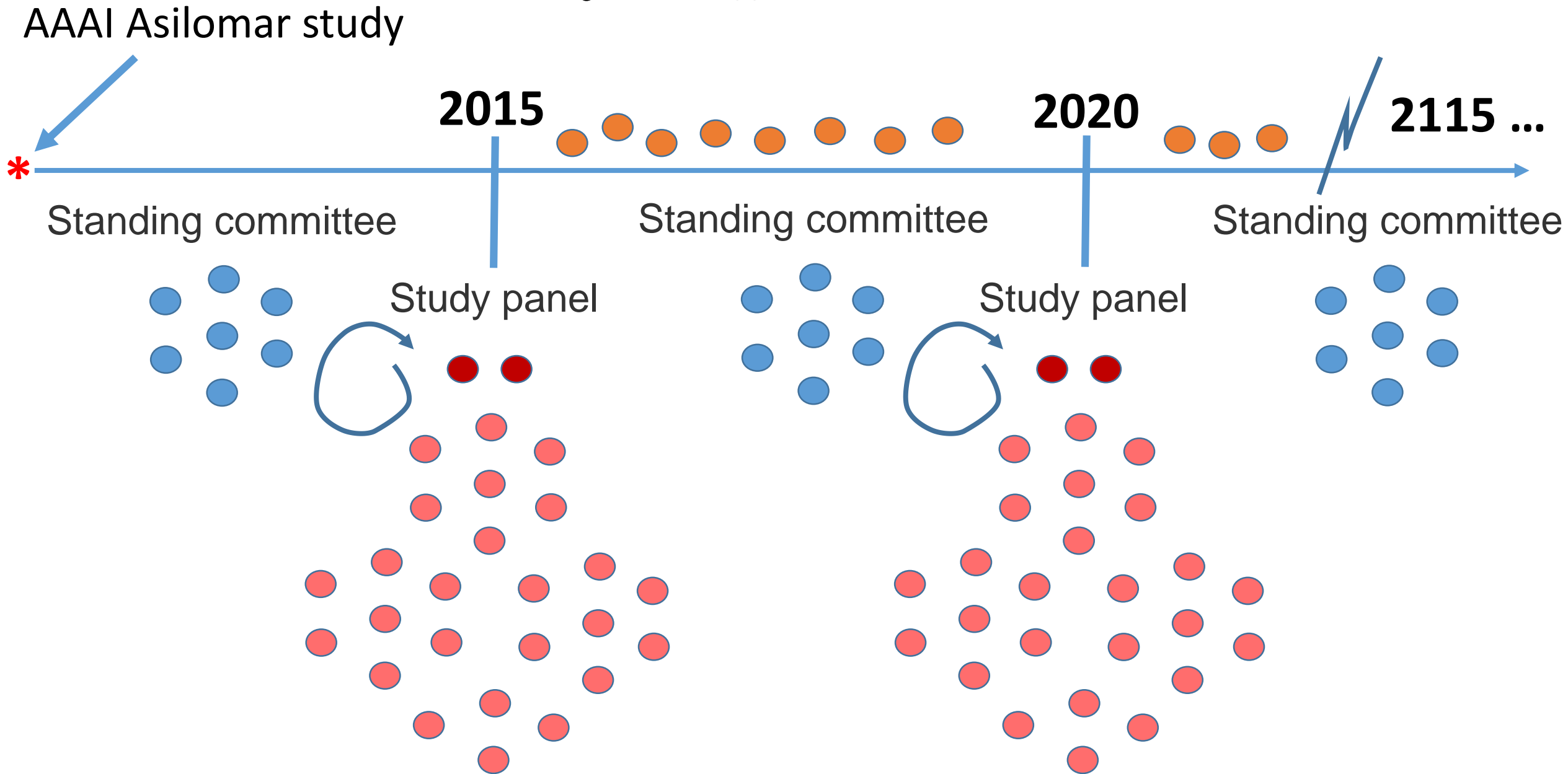
Timeline Backbone Close-Up

AAAI Asilomar study

Stanford Digital Archive

2015

Standing committee

Study panel

Convey results to multiple audiences

AI researchers

General public

Industry

Policy makers
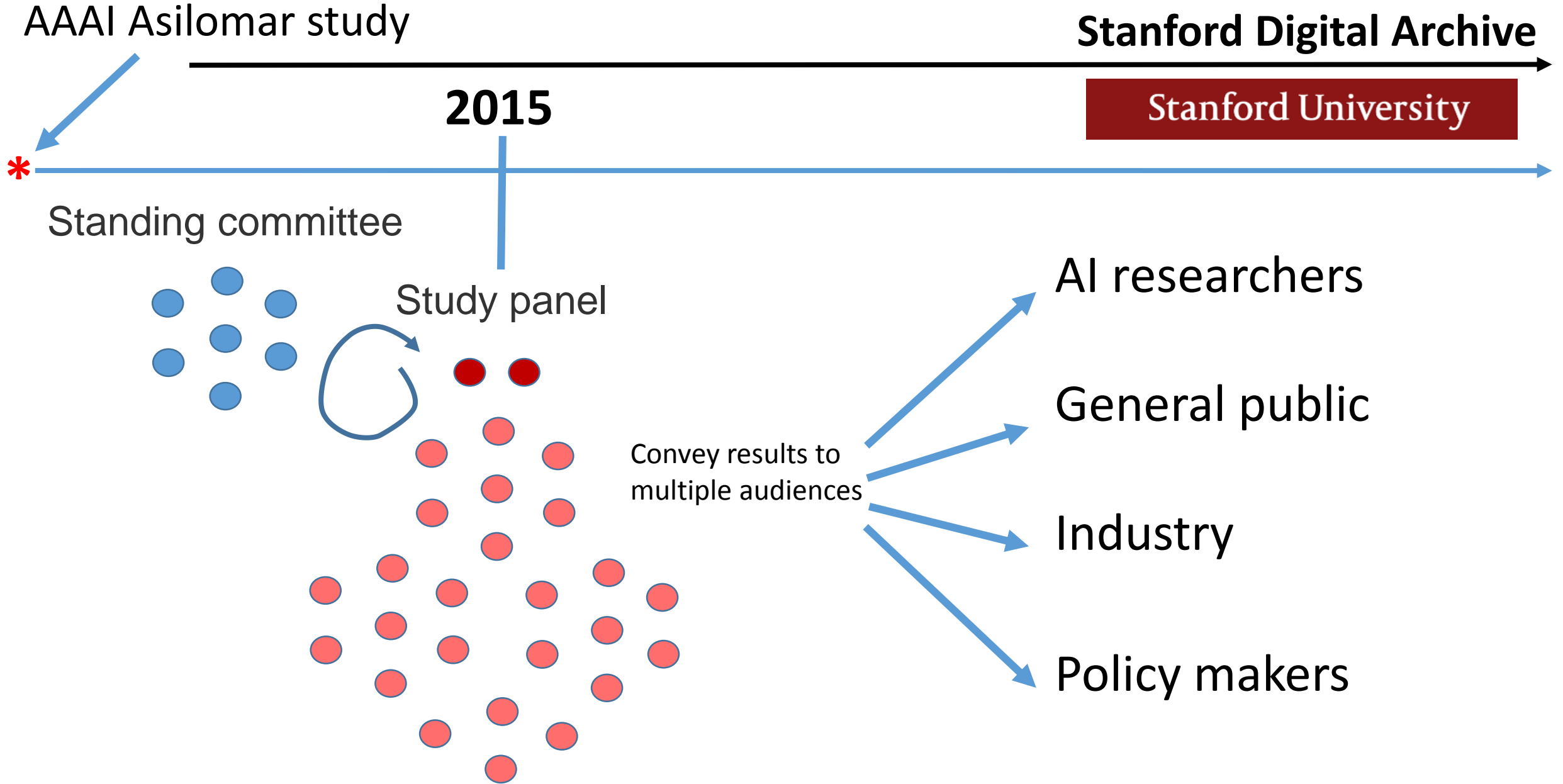
Designing Study Panels: Perspectives and Challenges

- Current state, future potential, possible surprises, consequences
- Technological advances in the short-term
- Science that could enable advances in the long-term
- Domains for framing issues and focus, such as:
  - Healthcare
  - Education
  - Work
  - Sustainability
  - Transportation
  - Governing, democracy and freedom
  - Public safety and defense

AI 100: IJCAI Exercise:
Choose a Real World Context And . . .

- Group 1: Define an AI capability to enhance health care or sustainability or transportation that should happen *now* and one barrier that stands in the way of its adoption.   **GREEN**

- Group 2: Identify a (new) opportunity for AI to enhance the world 5-10 years from now and 1-2 scientific advances needed for it.   **PURPLE**

- Group 3: Identify 2-3 core strengths of current AI and some way to use them to societal advantage.   **BLUE**

- Group 4: Identify 2-3 weaknesses of current AI and some approaches to overcoming them.   **YELLOW**

- Group 5: Identify a potential costly (or catastrophic) outcome of deployed AI systems and ways the field could proactively diminish its possibility.   **PINK**

- Group 6: Identify a major intellectual challenge in AI that must be met to meet a core problem in your chosen real world context.   **WHITE**

Your input and involvement is invited:

More information: http://ai100.stanford.edu

Email:
- ideas for study panels:
  ai100-studypanel@lists.stanford.edu
- general inquiries:
  ai100-info@lists.stanford.edu

One Hundred Year Study