Partnership of AI (2016): https://www.partnershiponai.org/tenets/

We are on a mission to shape best practices, research, and public dialogue about AI's benefits for people and society. Our members range from for-profit technology companies to representatives of civil society, to academic and research institutions, to start-ups and beyond.

Partners: Access Now, AI4ALL, AAAi, Accenture, ACLU, Affectiva, AI Forum New Zealand, AI Now Institute, Alan Turing Institute, Allen Instutute for Artificial Intelligence (AI2), Amazon, American Psychological Association, Amnesty International, Apple, |||||Australian National University, HCRI, The Center for Internet Society, Data&Society, T..., Catapult Digital, Digitial Asia, EFF, Fraunhofer, Firstdraft, Future of Humanity Instituten Future of Life, G3ict, The Hong Kong University of science and technology, HRDAG, Insight, The Joint Center for political and economic studies, CFI, Santa , Clara University, MIT Digital, Mozzila Foundation, Generation Artificial Intelligence Research Center, OPTIC, Salasource, Tufts, UL, UIL, UNDP, VIPlab

Tenets

We believe that artificial intelligence technologies hold great promise for raising the quality of people's lives and can be leveraged to help humanity address important global challenges such as climate change, food, inequality, health, and education.

Our members believe in and endeavor to uphold the following tenets:

1

We will seek to ensure that AI technologies benefit and empower as many people as possible.

2

We will educate and listen to the public and actively engage stakeholders to seek their feedback on our focus, inform them of our work, and address their questions.

3

We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI.

4

We believe that AI research and development efforts need to be actively engaged with and accountable to a broad range of stakeholders.

5

We will engage with and have representation from stakeholders in the business community to help ensure that domain-specific concerns and opportunities are understood and addressed.

6

We will work to maximize the benefits and address the potential challenges of AI technologies, by:

1. Working to protect the privacy and security of individuals.
2. Striving to understand and respect the interests of all parties that may be impacted by AI advances.
3. Working to ensure that AI research and engineering communities remain socially responsible, sensitive, and engaged directly with the potential influences of AI technologies on wider society.
4. Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.
5. Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.

7

We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.

8

We strive to create a culture of cooperation, trust, and openness among AI scientists and engineers to help us all better achieve these goals.

Our Goals

1

Develop and share best practices

Support research, discussions, identification, sharing, and recommendation of best practices in the research, development, testing, and fielding of AI technologies. Address such areas as fairness and inclusivity, explanation and transparency, security and privacy, values and ethics, collaboration between people and AI systems, interoperability of systems, and of the trustworthiness, reliability, containment, safety, and robustness of the technology.

2

Advance public understanding

Advance wider public understanding and awareness of AI by sharing insights into AI’s core technologies, potential benefits – and costs. We will act as a trusted experts on AI for society and their leaders, and will work to increase public understanding of how AI is progressing.

3

Provide an open and inclusive platform for discussion & engagement

Create and support opportunities for AI researchers and key stakeholders, including people in technology, law, policy, government, civil liberties, and the greater public, to communicate directly and openly with each other about relevant issues to AI and its influences on people and society. Ensure that key stakeholders have the knowledge, resources, and overall capacity to participate fully.

4

Identify and foster aspirational efforts in AI for socially beneficial purposes

Seek out, support, celebrate, and highlight aspirational efforts in AI for socially benevolent applications. Identify areas of untapped opportunity, including promising technologies and applications not being explored by academia and industry R&D.

Our Work (Thematic Pillars)

- 1

Safety-Critical AI

Advances in AI have the potential to improve outcomes, enhance quality, and reduce costs in such safety-critical areas as healthcare and transportation. Effective and careful applications of pattern recognition, automated decision making, and robotic systems show promise for enhancing the quality of life and preventing thousands of needless deaths.

However, where AI tools are used to supplement or replace human decision-making, we must be sure that they are safe, trustworthy, and aligned with the ethics and preferences of people who are influenced by their actions.

We will pursue studies and best practices around the fielding of AI in safety-critical application areas.

- 2

Fair, Transparent, and Accountable AI

AI has the potential to provide societal value by recognizing patterns and drawing inferences from large amounts of data. Data can be harnessed to develop useful diagnostic systems and recommendation engines, and to support people in making breakthroughs in such areas as biomedicine, public health, safety, criminal justice, education, and sustainability.

While such results promise to provide real benefits, we need to be sensitive to the possibility that there are hidden assumptions and biases in data, and therefore in the systems built from that data — in addition to a wide range of other system choices which can be impacted by biases, assumptions, and limits. This can lead to actions and recommendations that replicate those biases, and have serious blind spots.

Researchers, officials, and the public should be sensitive to these possibilities and we should seek to develop methods that detect and correct those errors and biases, not replicate them. We also need to work to develop systems that can explain the rationale for inferences.

We will pursue opportunities to develop best practices around the development and fielding of fair, explainable, and accountable AI systems.

- 3

AI, Labor, and the Economy

AI advances will undoubtedly have multiple influences on the distribution of jobs and nature of work. While advances promise to inject great value into the economy, they can also be the source of disruptions as new kinds of work are created and other types of work become less needed due to automation.

Discussions are rising on the best approaches to minimizing potential disruptions, making sure that the fruits of AI advances are widely shared and competition and innovation are encouraged and not stifled. We seek to study and understand best paths forward, and play a role in this discussion.

- 4

Collaborations Between People and AI Systems

A promising area of AI is the design of systems that augment the perception, cognition, and problem-solving abilities of people. Examples include the use of AI technologies to help physicians make more timely and accurate diagnoses and assistance provided to drivers of cars to help them to avoid dangerous situations and crashes.

Opportunities for R&D and for the development of best practices on AI-human collaboration include methods that provide people with clarity about the understandings and confidence that AI systems have about situations, means for coordinating human and AI contributions to problem solving, and enabling AI systems to work with people to resolve uncertainties about human goals.

- 5

Social and Societal Influences of AI

AI advances will touch people and society in numerous ways, including potential influences on privacy, democracy, criminal justice, and human rights. For example, while technologies that personalize information and that assist people with recommendations can provide people with valuable assistance, they could also inadvertently or deliberately manipulate people and influence opinions.

We seek to promote thoughtful collaboration and open dialogue about the potential subtle and salient influences of AI on people and society.

- 6

AI and Social Good

AI offers great potential for promoting the public good, for example in the realms of education, housing, public health, and sustainability. We see great value in collaborating with public and private organizations, including academia, scientific societies, NGOs, social entrepreneurs, and interested private citizens to promote discussions and catalyze efforts to address society’s most pressing challenges.

Some of these projects may address deep societal challenges and will be moonshots – ambitious big bets that could have far-reaching impacts. Others may be creative ideas that could quickly produce positive results by harnessing AI advances.