

Beijing Academy of Artificial Intelligence

Beijing AI Principles

On May 28, 2019, the Beijing Academy of Artificial Intelligence (BAAI) released the “Beijing AI Principles,” an outline to guide the research and development, implementation, and governance of AI.

The development of Artificial Intelligence (AI) concerns the future of the whole society, all humankind, and the environment. The principles below are proposed as an initiative for the research, development, use, governance and long-term planning of AI, calling for its healthy development to support the construction of a human community with a shared future, and the realization of beneficial AI for humankind and nature.

Research and Development

The research and development (R&D) of AI should observe the following principles:

- ◆ **Do Good:** AI should be designed and developed to promote the progress of society and human civilization, to promote the sustainable development of nature and society, to benefit all humankind and the environment, and to enhance the well-being of society and ecology.
- ◆ **For Humanity:** The R&D of AI should serve humanity and conform to human values as well as the overall interests of humankind. Human privacy, dignity, freedom, autonomy, and rights should be sufficiently respected. AI should not be used to against, utilize or harm human beings.
- ◆ **Be Responsible:** Researchers and developers of AI should have sufficient considerations for the potential ethical, legal, and social impacts and risks brought in by their products and take concrete actions to reduce and avoid them.
- ◆ **Control Risks:** Continuous efforts should be made to improve the maturity, robustness, reliability, and controllability of AI systems, so as to ensure the security for the data, the safety and security for the AI system itself, and the safety for the external environment where the AI system deploys.
- ◆ **Be Ethical:** AI R&D should take ethical design approaches to make the system trustworthy. This may include, but not limited to: making the system as fair as possible, reducing possible discrimination and biases, improving its transparency, explainability, and predictability, and making the system more traceable, auditable and accountable.
- ◆ **Be Diverse and Inclusive:** The development of AI should reflect diversity and inclusiveness, and be designed to benefit as many people as possible, especially those who would otherwise be easily neglected or underrepresented in AI applications.
- ◆ **Open and Share:** It is encouraged to establish AI open platforms to avoid data/platform monopolies, to share the benefits of AI development to the greatest extent, and to promote equal development opportunities for different regions and industries. Use

The use of AI should observe the following principles:

- ◆ **Use Wisely and Properly:** Users of AI systems should have the necessary knowledge and ability to make the system operate ac-

ording to its design, and have sufficient understanding of the potential impacts to avoid possible misuse and abuse, so as to maximize its benefits and minimize the risks.

- ◆ **Informed-consent:** Measures should be taken to ensure that stakeholders of AI systems are with sufficient informed-consent about the impact of the system on their rights and interests. When unexpected circumstances occur, reasonable data and service revocation mechanisms should be established to ensure that users’ own rights and interests are not infringed.
- ◆ **Education and Training:** Stakeholders of AI systems should be able to receive education and training to help them adapt to the impact of AI development in psychological, emotional and technical aspects.

The governance of AI should observe the following principles:

- ◆ **Optimizing Employment:** An inclusive attitude should be taken towards the potential impact of AI on human employment. A cautious attitude should be taken towards the promotion of AI applications that may have huge impacts on human employment. Explorations on Human-AI coordination and new forms of work that would give full play to human advantages and characteristics should be encouraged.
- ◆ **Harmony and Cooperation:** Cooperation should be actively developed to establish an interdisciplinary, cross-domain, cross-sectoral, cross-organizational, cross-regional, global and comprehensive AI governance ecosystem, so as to avoid malicious AI race, to share AI governance experience, and to jointly cope with the impact of AI with the philosophy of “Optimizing Symbiosis”.

Adaptation and Moderation:

Adaptive revisions of AI principles, policies, and regulations should be actively considered to adjust them to the development of AI. Governance measures of AI should match its development status, not only to avoid hindering its proper utilization, but also to ensure that it is beneficial to society and nature.

Subdivision and Implementation:

Various fields and scenarios of AI applications should be actively considered for further formulating more specific and detailed guidelines. The implementation of such principles should also be actively promoted – through the whole life cycle of AI research, development, and application.

Long-term Planning:

Continuous research on the potential risks of Augmented Intelligence, Artificial General Intelligence (AGI) and Superintelligence should be encouraged. Strategic designs should be considered to ensure that AI will always be beneficial to society