

The Asilomar AI Principles

- Available at: <https://futureoflife.org/ai-principles/>
- 23 principles, arranged across 3 subheadings (Research Issues, Ethics and Values, and Longer-Term Issues)
- Widely-endorsed and cited as influential to other Principles, policy development.
- Principles cover a range of issues from the need for a science-policy link (“There should be constructive and healthy exchange between AI researchers and policy-makers.”) to issues of the technical function of AI systems (“Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.”)

Selected AI Principles

13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

17) Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

20) Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.

21) Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.