

Link original: <https://futureoflife.org/open-letter/ai-principles/>

Princípios de Asilomar para IA

Os Princípios de Asilomar para IA, coordenados pelo FLI e desenvolvidos na conferência Beneficial AI 2017, são um dos primeiros e mais influentes conjuntos de princípios de governança de IA.

Esses princípios foram desenvolvidos em conjunto com a Conferência de Asilomar (<https://futureoflife.org/event/bai-2017/>) de 2017 (vídeos aqui: <https://www.youtube.com/playlist?list=PLpxRpA6hBNrwA8DlvNyIOO9B97wADE1t>), através do processo descrito aqui (<https://futureoflife.org/principles/principled-ai-discussion-asilomar/>).

A inteligência artificial já forneceu ferramentas benéficas que são usadas todos os dias por pessoas ao redor do mundo. Seu desenvolvimento contínuo, guiado pelos seguintes princípios, oferecerá oportunidades incríveis para ajudar e capacitar as pessoas nas próximas décadas e séculos.

Questões de Pesquisa

1. **Objetivo da Pesquisa:** O objetivo da pesquisa em IA deve ser criar inteligência benéfica, não inteligência não direcionada.
2. **Financiamento da Pesquisa:** Os investimentos em IA devem ser acompanhados por financiamento para pesquisas sobre a garantia de seu uso benéfico, incluindo questões complexas em ciência da computação, economia, direito, ética e estudos sociais, tais como: Como podemos tornar os sistemas de IA futuros altamente robustos, para que façam o que queremos sem falhar ou serem hackeados? Como podemos aumentar nossa prosperidade através da automação enquanto mantemos os recursos e o propósito das pessoas? Como podemos atualizar nossos sistemas legais para serem mais justos e eficientes, acompanhar o ritmo da IA e gerenciar os riscos associados à IA? Com quais valores a IA deve estar alinhada e qual status legal e ético ela deve ter?
3. **Vínculo Ciência-Política:** Deve haver uma troca construtiva e saudável entre pesquisadores de IA e formuladores de políticas.
4. **Cultura de Pesquisa:** Deve ser fomentada uma cultura de cooperação, confiança e transparência entre pesquisadores e desenvolvedores de IA.
5. **Evitar a Corrida:** As equipes que desenvolvem sistemas de IA devem cooperar ativamente para evitar a negligência nos padrões de segurança

Ética e Valores

- 6) **Segurança:** Os sistemas de IA devem ser seguros durante toda a sua vida operacional, e verificavelmente seguros onde aplicável e viável.
7. **Transparência em Caso de Falhas:** Se um sistema de IA causar danos, deve ser possível determinar o porquê.

8. **Transparência Judicial:** Qualquer envolvimento de um sistema autônomo na tomada de decisões judiciais deve fornecer uma explicação satisfatória auditável por uma autoridade humana competente.
9. **Responsabilidade:** Os projetistas e construtores de sistemas avançados de IA são partes interessadas nas implicações morais de seu uso, mau uso e ações, com uma responsabilidade e oportunidade de moldar essas implicações.
10. **Alinhamento de Valores:** Sistemas de IA altamente autônomos devem ser projetados de modo que seus objetivos e comportamentos possam ser garantidos a alinhar-se com os valores humanos durante toda a sua operação.
11. **Valores Humanos:** Os sistemas de IA devem ser projetados e operados de forma a serem compatíveis com os ideais de dignidade humana, direitos, liberdades e diversidade cultural.
12. **Privacidade Pessoal:** As pessoas devem ter o direito de acessar, gerenciar e controlar os dados que geram, dada a capacidade dos sistemas de IA de analisar e utilizar esses dados.
13. **Liberdade e Privacidade:** A aplicação da IA aos dados pessoais não deve limitar de maneira irracional a liberdade real ou percebida das pessoas.
14. **Benefício Compartilhado:** As tecnologias de IA devem beneficiar e capacitar o maior número possível de pessoas.
15. **Prosperidade Compartilhada:** A prosperidade econômica criada pela IA deve ser amplamente compartilhada, para beneficiar toda a humanidade.
16. **Controle Humano:** Os humanos devem escolher como e se delegam decisões aos sistemas de IA, para alcançar objetivos escolhidos pelos humanos.
17. **Não Subversão:** O poder conferido pelo controle de sistemas de IA altamente avançados deve respeitar e melhorar, em vez de subverter, os processos sociais e cívicos dos quais depende a saúde da sociedade.

18. **Corrida Armamentista de IA:** Deve-se evitar uma corrida armamentista em armas autônomas letais.

Questões de Longo Prazo

- 19) Cautela com Capacidades: Não havendo consenso, devemos evitar suposições fortes sobre os limites superiores das capacidades futuras da IA.

20. Importância: A IA avançada pode representar uma mudança profunda na história da vida na Terra e deve ser planejada e gerida com cuidado e recursos proporcionais.