# Recursos Bibliográficos

# Epistemologia e Filosofia da Mente

### Hubert Dreyfus

Dreyfus é um filósofo americano conhecido por sua crítica pioneira à Inteligência Artificial. Ele argumenta que a inteligência humana e a expertise dependem principalmente de processos inconscientes que não podem ser completamente capturados em regras formais. [Wikipedia](https://en.wikipedia.org/wiki/Hubert_Dreyfus's_views_on_artificial_intelligence) Sua crítica se baseia nas perspectivas de filósofos como Merleau-Ponty e Heidegger.

Obras principais:

* "What Computers Can't Do: The Limits of Artificial Intelligence" (1972)
* "Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer" (1986)
* "What Computers Still Can't Do" (1992)

Link relevante:<https://en.wikipedia.org/wiki/Hubert_Dreyfus's_views_on_artificial_intelligence>

### Daniel Dennett

Dennett é um filósofo americano cujo trabalho se concentra na filosofia da mente e da ciência cognitiva. Sua pesquisa se centra na filosofia da mente, filosofia da ciência e filosofia da biologia, particularmente como esses campos se relacionam com a biologia evolutiva e a ciência cognitiva. [Wikipedia](https://en.wikipedia.org/wiki/Daniel_Dennett)

Obras principais:

* "Consciousness Explained" (1991)
* "Darwin's Dangerous Idea" (1995)
* "From Bacteria to Bach and Back" (2017)

Link relevante:<https://now.tufts.edu/2023/10/02/daniel-dennetts-been-thinking-about-thinking-and-ai>

### John Searle

Searle é conhecido pelo seu famoso argumento do "Quarto Chinês". Ele argumenta que o experimento mental destaca o fato de que computadores apenas usam regras sintáticas para manipular símbolos, mas não têm compreensão de significado ou semântica. [Stanford](https://plato.stanford.edu/entries/chinese-room/)

Obras principais:

* "Minds, Brains, and Programs" (1980)
* "The Rediscovery of the Mind" (1992)

Link relevante:<https://plato.stanford.edu/entries/chinese-room/>

## Ética Aplicada e Filosofia da Técnica

### Hans Jonas

Jonas é conhecido por seu trabalho sobre ética da tecnologia. Ele argumentou que a tecnologia requer uma ética na qual a responsabilidade é o imperativo central, pois pela primeira vez na história somos capazes de destruir a Terra e a humanidade. [Stanford](https://plato.stanford.edu/entries/technology/)

Obras principais:

* "The Imperative of Responsibility" (1979)
* "The Phenomenon of Life: Toward a Philosophical Biology" (1966)

Link relevante:<https://en.wikipedia.org/wiki/Hans_Jonas>

### Luciano Floridi

Floridi é um filósofo italiano conhecido por seu trabalho em filosofia e ética da informação. Sua análise filosófica das novas tecnologias leva a um novo quadro metafísico no qual nossa compreensão da natureza última da realidade muda de uma materialista para uma informacional. [SpringerLink](https://link.springer.com/book/10.1007/978-94-007-4292-5)

Obras principais:

* "The Ethics of Information" (2013)
* "The Fourth Revolution: How the Infosphere is Reshaping Human Reality" (2014)

### Byung-Chul Han

Han é um filósofo e teórico cultural sul-coreano que vive na Alemanha. Seu trabalho se concentra principalmente em uma crítica ao neoliberalismo e seu impacto na sociedade e no indivíduo. [Wikipedia](https://en.wikipedia.org/wiki/Byung-Chul_Han)

Obras principais:

* "The Burnout Society" (2015)
* "Psychopolitics: Neoliberalism and New Technologies of Power" (2017)
* "The Transparency Society" (2015)

Link relevante:<https://newintrigue.com/2020/06/29/the-philosophy-of-byung-chul-han/>

## IA e Sociedade

### Nick Bostrom

Bostrom é filósofo sueco-britânico especializado em riscos existenciais e superinteligência. Bostrom acredita que avanços em inteligência artificial podem levar à superinteligência, que ele define como "qualquer intelecto que exceda grandemente o desempenho cognitivo dos humanos em praticamente todos os domínios de interesse". [Wikipedia](https://en.wikipedia.org/wiki/Nick_Bostrom)

Obras principais:

* "Superintelligence: Paths, Dangers, Strategies" (2014)
* "Anthropic Bias: Observation Selection Effects in Science and Philosophy" (2002)
* "Deep Utopia: Life and Meaning in a Solved World" (2024)

Link relevante:<https://nickbostrom.com/>

### Shoshana Zuboff

Zuboff é professora emérita da Harvard Business School e autora do influente livro "The Age of Surveillance Capitalism". Seu livro é uma espécie de reencarnação do foco de Wiener na computação como controle, explorando como as grandes empresas de tecnologia utilizam nossos dados para prever e modificar comportamentos. [Expensivity](https://www.expensivity.com/erik-j-larson-reviews-surveillance-capitalism/)

Obras principais:

* "The Age of Surveillance Capitalism" (2019)

### Cathy O'Neil

O'Neil é matemática e autora que explora como algoritmos e modelos matemáticos podem perpetuar desigualdades existentes. Seu trabalho é frequentemente citado em discussões sobre ética de dados e algoritmos.

Obras principais:

* "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy" (2016)

## Interpretabilidade em IA e Perspectivas Recentes

### Dario Amodei

Amodei é CEO da Anthropic e pesquisador em IA. Em seus escritos recentes, ele discute como a interpretabilidade em IA pode se aplicar à neurociência, sugerindo que o mecanismo computacional descoberto por pesquisadores de interpretabilidade em sistemas de IA foi recentemente redescoberto nos cérebros de ratos. [Darioamodei](https://darioamodei.com/machines-of-loving-grace)

Obras relevantes:

* "Machines of Loving Grace" (ensaio, 2024)

Link relevante:<https://darioamodei.com/machines-of-loving-grace>

# Research: Do LLMs Have Values?

by [Jordan Loewen-Colón](https://hbr.org/search?term=Jordan%20Loewen-Col%C3%B3n), [Benedict Heblich](https://hbr.org/search?term=Benedict%20Heblich) and [Marius Birkenbach](https://hbr.org/search?term=Marius%20Birkenbach)

Link:

[https://hbr.org/2025/05/research-do-llms-have-values?tpcc=whatsapp\_edit&utm\_campaign=hbr&utm\_medium=social&utm](https://hbr.org/2025/05/research-do-llms-have-values?tpcc=whatsapp_edit&utm_campaign=hbr&utm_medium=social&utm_source=whatsapp)