

# 超知能のある社会

エンジニアリングデー・シンポジウム

2025-03-04

丸山 宏

花王/東大/PFN/EARTHBRAIN

X/Twitter: @maruyama

# 「人の心」はタマネギの皮のようなものか？

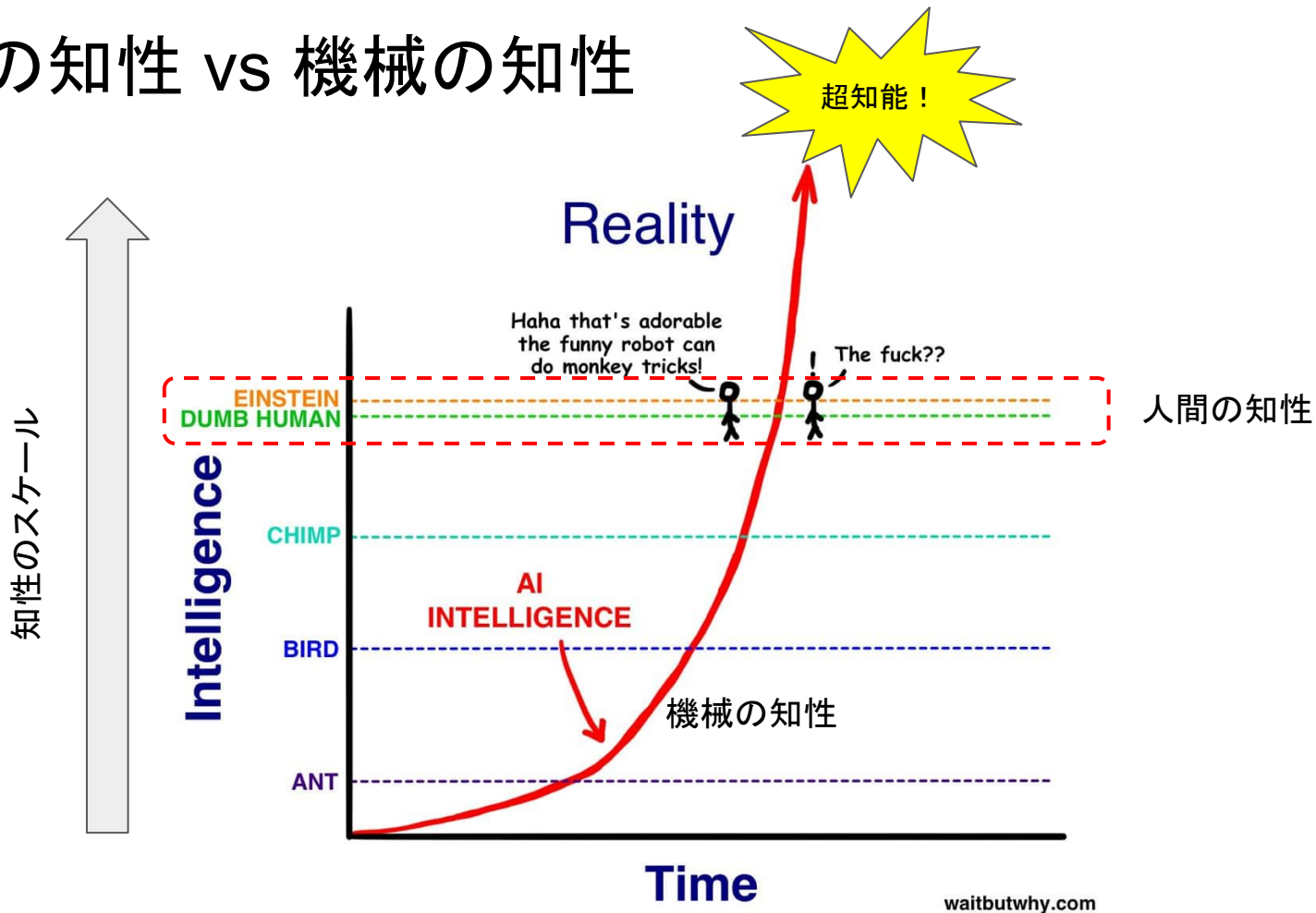
「たまねぎの皮」のアナロジーも助けになる。脳や心の機能を考えると、私たちが説明できるある動作は純粹に機械的なものであることに気づく。私たちは、これは本当の心ではなく、本当の心を見いだすために剥かなければならない皮のようなものだと言う。しかしどんなものを見つけようとも その先にはまだ剥くべき皮がある、以下同様。この手の進みかたで 私たちはほんとうに「真の」心にたどりつけるのだろうか、それとも ついにはもはや何も入っていない皮を剥いてしまうのだろうか？ 後者の場合、心は全てまるごと機械的だということになる。



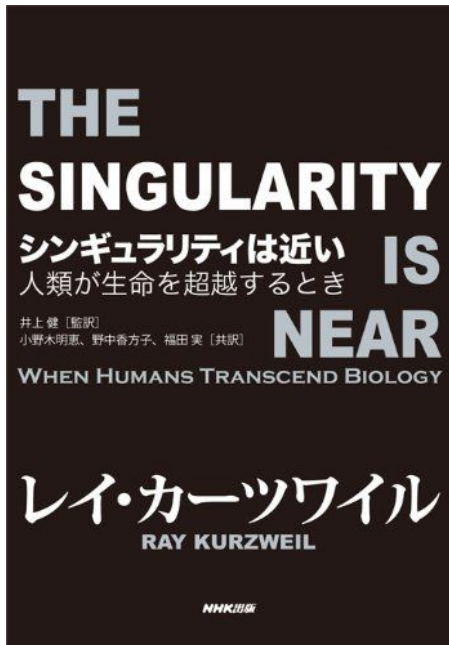
Alan Turing, Computing Machinery and Intelligence, 1950.

新山 祐介氏による翻訳、<http://www.unixuser.org/~euske/doc/turing-ja/index.html>

# 人の知性 vs 機械の知性



# レイ・カーツワイル「シンギュラリティは近い」



- エポック1: 宇宙の生成(物理定数の決定)
- エポック2: 生物の発生(DNAの情報)
- エポック3: 脳の進化 (脳による情報処理)
- エポック4: テクノロジ(情報処理ツール)
- エポック5: 技術と脳の融合(シンギュラリティ)
- エポック6: 宇宙の覚醒(宇宙全体の知性化)

シンギュラリティ:

“When humans transcend biology with the power of technology”  
(人類が技術によって生物の限界を超えるとき)

Singularity is here

(おそらく)

ChatGPTに関して、自民党AIプロジェクトチーム 安宅さんの資料より

# ユヴァル・ノア・ハラリによる「人類の立ち位置」

“人類は自然界の中でなぜ特別な存在なのか ”



自然界の中で平等(アニミズム)



農耕の発明

「神の子」としての特別な存在



科学の発展

「考える存在」としての優位性(ヒューマニズム)



情報技術

データの中に埋没する存在(「データイズム」)

あらゆる面で人間より知的に優れた超知性ができたとき、  
人類は自然界の中で特別な存在であることを止めるのか？

# 「AIによる脅威」の議論

【考察ノート】「AGIリスク」の議論にどう向き合えばいいのか

2024年9月29日 ver.1 公開

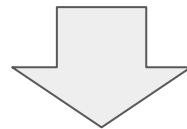
記：丸山隆一<sup>1</sup>

## はじめに

「AIによって引き起こされる絶滅リスクの軽減を、世界的な優先事項にすべきである」——2023年5月、米国のCenter for AI Safetyなる組織のウェブサイトに、こんなメッセージが掲げられた<sup>2</sup>。その下には、Geoffrey Hinton氏、Yoshua Bengio氏といった著名なAI研究者や、Demis Hassabis氏、Sam Altman氏といった最先端民間AIラボのリーダーたちの署名が連なる。しかし、この「AIによって引き起こされる絶滅リスク」という文言は、筆者にはやや奇異に響いた。

西洋の一部の界限では、AIによる「壊滅的リスク」「存亡リスク」ないし「長期リスク」をめぐる議論が存在している。このこと自体は広く知られている。しかし、そこで具体的に何が懸念されているのかについて、今のところ多くは伝わっていない。未来のリスクとして語られている内容が極めて不確実だし、そのリスクが実在するとしても、それに対して何ができるのかわかりにくい。どこまで真剣に扱えばよいのだろうか。筆者だけでなく、「AIによる絶滅リスク」と言われてもピンとこない

AIのコントロール喪失問題をはじめとするAGIリスクを気にする人が、白人男性に著しく偏っているのはなぜか



“強者のアジェンダ” だから

強者でない眼で超知能を見ると？？

[https://researchmap.jp/rmaruy/published\\_works](https://researchmap.jp/rmaruy/published_works)

# 私たちの将来: Contain(閉じ込め)か

Contain: 機械に制約を課し、人間に対する脅威とならないようにする

- EU「包括的AI規制法案」
- アシモフ「ロボット三原則」
- 漏れのないポリシーは定義不可能かもしれない
- ポリシーを守らない人が出てくるかもしれない
- 地球外文明が、人間中心主義を脅威とみなすかもしれない

# Embrace(抱擁)か

Embrace: 機械との一体化により、人類文明全体として高次の存在を目指す

- カーツワイル「シンギュラリティ」
- A.C.クラーク「幼年期の終わり」
- 個人のアイデンティティは曖昧になるかもしれない
- いろいろな意味で「人間らしさ」を失うかもしれない
- 「基本的人権」を見直す必要があるかもしれない

Thank You

@maruyama on X/twitter



ISBN-13 : 978-4764906068

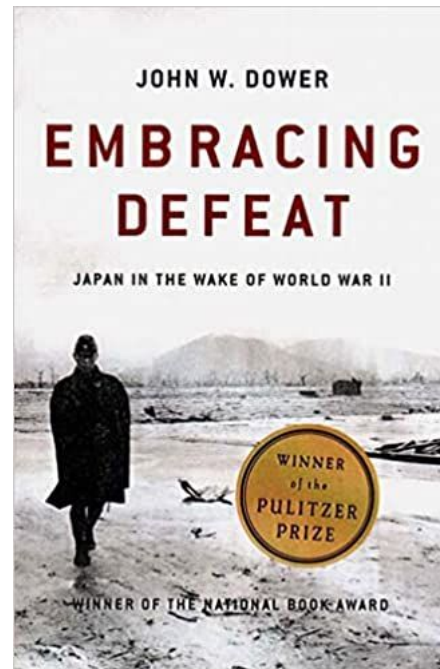


以下バックアップ

# 1945年8月～ それまで敵だった欧米の価値観を抱擁した日本



ジョン・ダワー：1938年生まれ。アメリカ合衆国の歴史学者。マサチューセッツ工科大学名誉教授。専攻は、日本近代史。1958年来日。妻は日本人。米国における日本占領研究の第一人者であり、1970年代の日本近代化論の批判でも知られる。



John Dower, *Embracing Defeat — Japan in the Wake of World War II*, ISBN 978-0393320275, 2000.

## エンジニアの役割:「ハクティビスト」

ユヴァル・ノア・ハラリとオードリー・タンの対談

# TO BE OR NOT TO BE HACKED?

The future of democracy, work, and identity  
Yuval Noah Harari in conversation with Audrey Tang  
Moderated by Puja Ohlhaber

RADICALXCHANGE

タン: 市井の“Hacktivist” が世の中を常にモニタしている世界