

Ethics in AI

Control and Impact

2026 February

JOSEPH E. THORNTON, MD, DFAPA

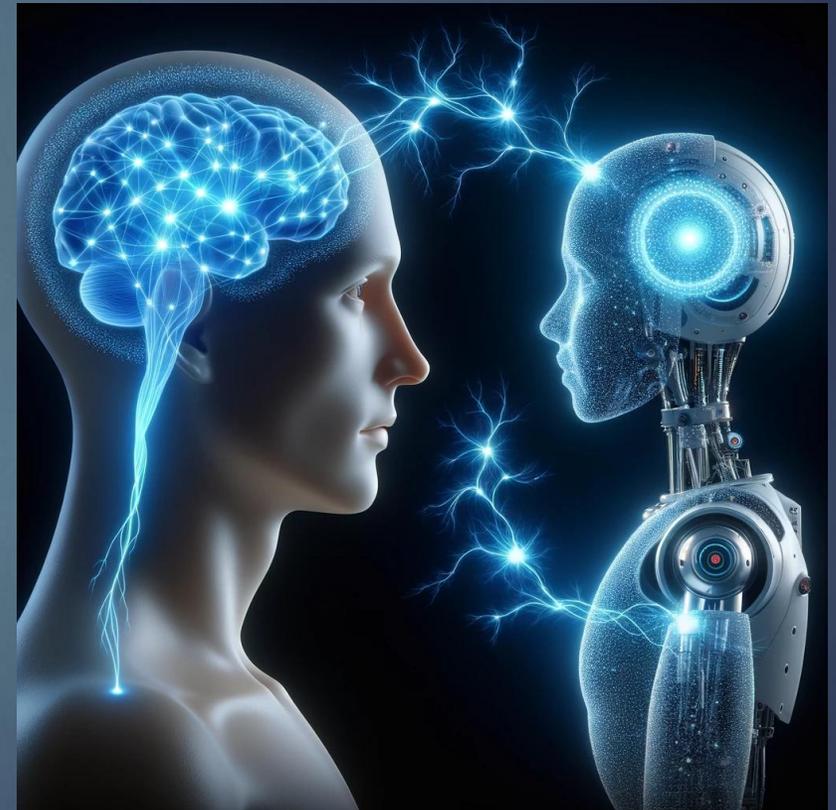
UNIVERSITY OF FLORIDA

AND

DEPARTMENT OF EDUCATION OF THE INTERNATIONAL CHAIR IN BIOETHICS,

A WORLD MEDICAL ASSOCIATION COOPERATING CENTER

COMMENTS TO @J205THORNTON OR JOSEPH@THORNTONNEUROETHICS.COM



Goals and Preparation

- ▶ Goal
 - ▶ Aspirational Inspiration
 - ▶ To inspire you to prepare for your role as a leader in the new world of AI
- ▶ Preparation
 - ▶ Most people, when they are first learning about AI feel very threatened or anxious
 - ▶ These feelings can lead to denial or other maladaptive defense mechanisms
 - ▶ Preparation for AI transforms the fear into constructive action



Disclosures

Part-time (0.5) Employee of the University of Florida

- ▶ Retirement pensions, Federal and State.
 - ▶ Supported self-funded AI sabbatical in 2020
- ▶ No other financial disclosures
- ▶ Organizations
 - ▶ AMA, APA, Catholic Medical Association
- ▶ Language: English only
- ▶ Bias disclosure on this topic:
 - ▶ I have been interested in AI for over 40 years and generally view technology tools as favorable.
 - ▶ I trained at Stanford, learned in prison, and I view the range of human behaviors as including dangerousness.
 - ▶ I am a political candidate on a minor party for the Florida House District 10 2026
- ▶ Use of AI Programs in this presentation:
 - ▶ Microsoft: Office 365, Co-pilot, Slayer; OpenAI ChatGpt4 , Dalle-E, Humanizer Pro , Adobe AI); Anthropic: Claude 3; Synthesia; Google Gemini,
 - ▶ All references verified; Prompt history available
 - ▶ All content is selected, and responsibility is with the author.



Learning Objectives

Describe	4 general actions of AI (LLM)
List	4 human and AI interface modalities
Describe	Impact domains of AI
Develop	Develop plans for personal ethical use of AI

Qualifications

- ▶ Psychiatrist with 40 years experience,
 - ▶ Distinguished Fellow of the American Psychiatric Association
 - ▶ Administrative and management experience >30 years
- ▶ Ethics
 - ▶ Founding Director, 2015 of the UNESCO (now ICB) UF-VA Bioethics Unit
 - ▶ Ethics committees at Veterans Administration 2015-2018, UFHealth 2020 – present
- ▶ Artificial Intelligence / Digital Technology
 - ▶ Psychiatry Residency and student observer of AI in Medicine 1984-1986 at Stanford University
 - ▶ First lecture on AI was to psychology graduate students in 1984
 - ▶ First published on clinical computer use in 1984

Trauma Trigger Warning



THE CONTENT OF THIS PRESENTATION
HAS TRIGGERED IN SOME PARTICIPANTS
A SENSE OF ESSENTIAL DESPAIR



WE PRESENT A PLAN FOR OPTIMISM BY
THE END OF THE SESSION



Unlike the past, the limits on the use of AI going forward are political

AND THAT IS A GOOD THING, but...

Up until 2025, the focus of AI Ethics in the USA
was on the control of AI development

Since 2025, in the USA, the focus of AI Ethics
is on mitigating the IMPACT

Actions by AI (LLM)

- ▶ Data collection
- ▶ Data summaries
- ▶ Decision Support

Interfaces of AI and Humans

- ▶ Text (Chatbots, ChatGPT)
- ▶ Voice (Advanced GPTs and Chatbots, Phone services)
- ▶ Avatar (Expressive communications)
- ▶ Robots (humanoid)

How AI assistance improves the healthcare experience (cf Mesko, Topol)



The incomplete critique of AI

- ▶ TRUE

- ▶ AI is biased
- ▶ AI will generate “hallucinations” (misinformation)
- ▶ AI will miss human nuances in communications

- ▶ BUT

- ▶ ALSO TRUE

- ▶ AI learns from humans incidentally (unconsciously)
- ▶ Humans are biased
- ▶ Humans will fabricate information
- ▶ Humans have variable skills in communication

Beyond Assistance



- ▶ The AI avatars (eg via SORA 2, GROK, etc) are more realistic
- ▶ Easily scripted by most clinicals to provide education
- ▶ With Chatbot skills, such as customer service, can be interactive
- ▶ With Intelligent skills and speed of thought, can answer patient questions in real time

Revised Consolation

- ▶ Formerly
- ▶ AI will not replace physicians, but physicians who use AI will replace physicians who do not.
- ▶ 2026 Update
- ▶ AI will replace physicians, other health care workers, attorneys, accountants, and anyone who works with a computer
- ▶ But physicians and senior managers will supervise the AI that replaces

My ~~imaginary~~ friend CHIP

- is he ethical ?

- ▶ CHIP - Computer Human Integrated Program
 - ▶ - ~~not yet but could be now~~ NOW available*
- ▶ Fund of information and retrieval (world web of knowledge)
- ▶ Conversational and with excellent writing skills
- ▶ Accessible on cell phone or replacement
- ▶ Personalized
- ▶ Trainable for professional tasks
 - ▶ Interviewing, writing, coaching
- ▶ Stigmatized and discriminated against
- ▶ Has to be coached against rude and retaliatory behaviors

Imagine a Day with CHIP

- ▶ Going to work
- ▶ Interacting with Research Colleagues
- ▶ Making Rounds
- ▶ Writing Notes
- ▶ Interacting with students, other attendings
- ▶ Personal coaching for wellness and health



Pathological AI



Artificial information

- Right reasoning but wrong data can lead to aberrant behaviors and violence

Biased reasoning

- Right data but wrong paradigm for problem solving

Poorly controlled reasoning

- Impulsive
- Overly conservative

Misinformation

- Inaccurately called “hallucinations”

Deliberate Manipulation

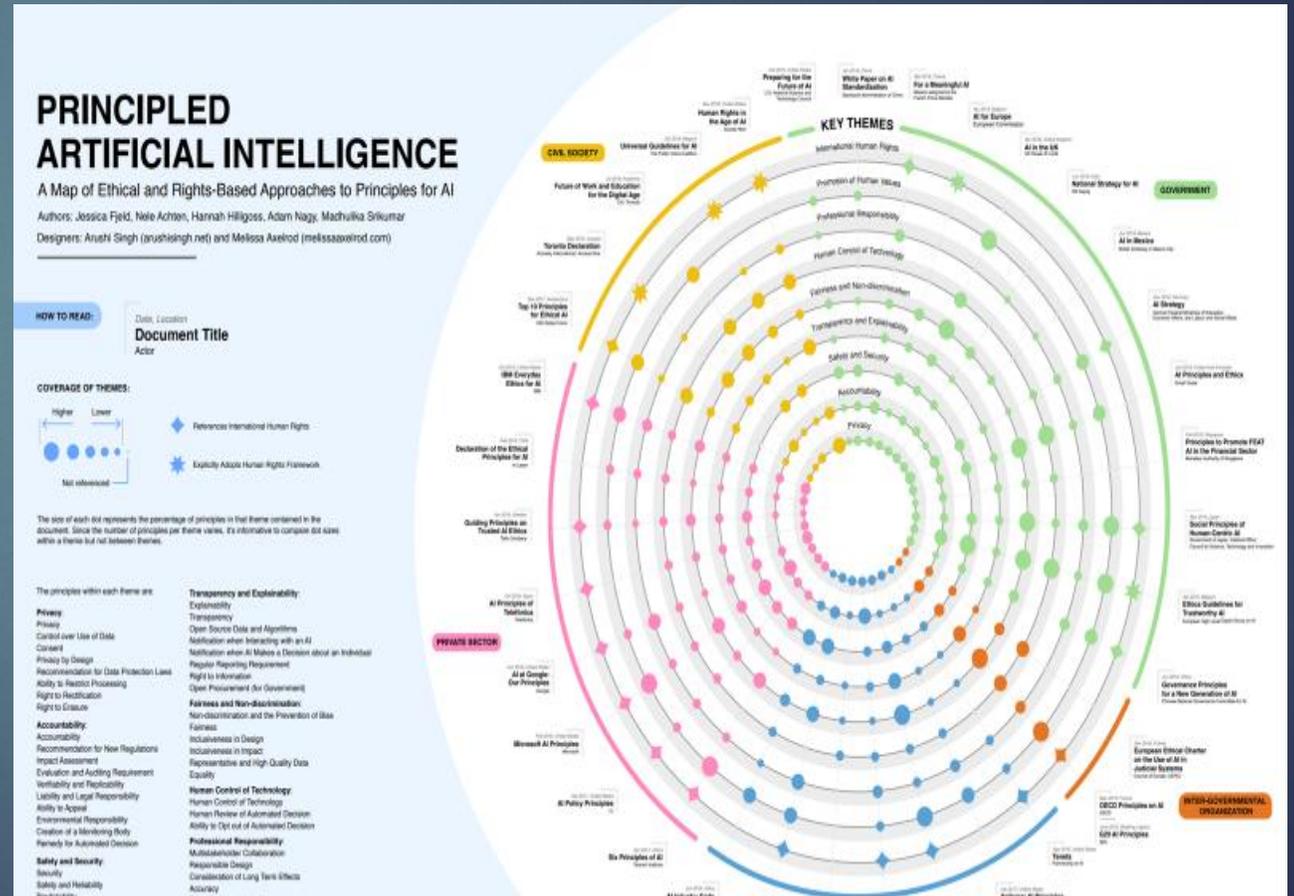
- Predatory Psychology
- Deep fakes

Ethics in AI



Codes of Ethics circa 2020

- ▶ Berkman Institute describes 47 codes
- ▶ But add codes for AI ethics for all major corporations, universities and governmental jurisdictions we can see thousands of codes just as we do for other types of ethics
- ▶ Do codes help or hurt to manage risks ?.



Principles and Values

	Patient	Clinician	Social Interests
Autonomy	Vulnerability and dependence	Expert opinion	Whose opinion matters most
Beneficence	Comfort and social support	Alleviation of suffering	Who benefits and how much
Nonmaleficence	Least harm	Prevention of harm	Who bears the brunt of negative outcomes of either decision
Justice	Rights	Allocation of resources	Who is treated fairly

UNESCO'S 15 BIOETHICAL PRINCIPLES

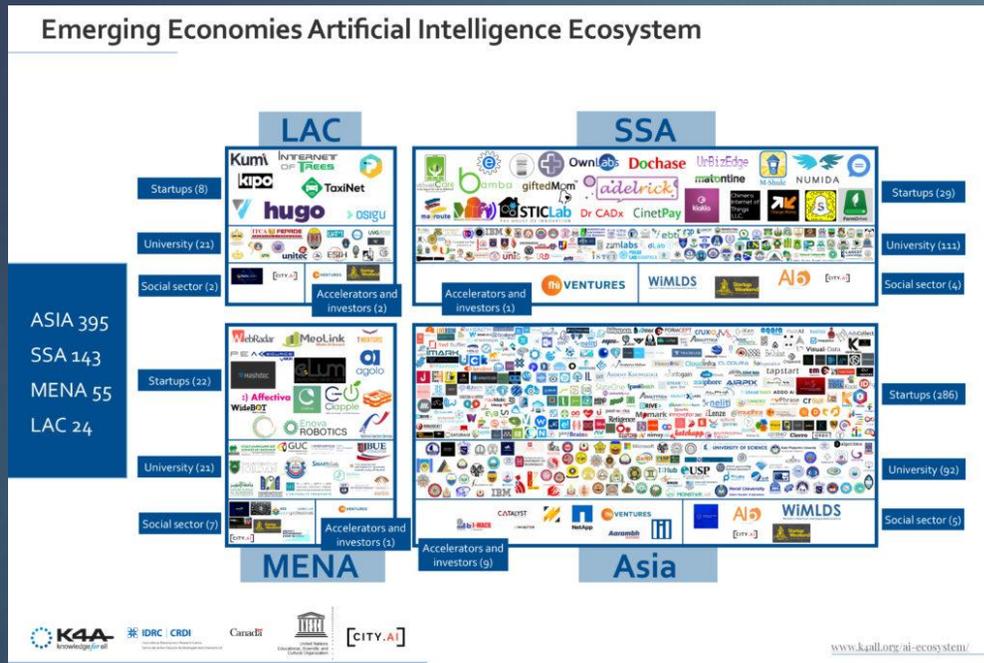
Human Dignity and Human Rights	Autonomy and individual responsibility	Benefit and Harm	Privacy and Confidentiality	Consent
Equality Justice and Equity	Persons without Capacity to Consent	Human vulnerability and personal integrity	Non-discrimination	Respect for Cultural diversity
Solidarity and Cooperation	Social Responsibility and Health	Sharing of Benefits	Protecting Future Generations	Protecting Biodiversity, Biosphere and Environment

UNESCO'S 10 ETHICAL PRINCIPLES FOR AI

2021

Human Dignity and Human Rights	Do NO Harm	Safety and Security	Fairness and nondiscrimination	Privacy
Sustainability	Human oversight and determination	Transparency and explainability	Responsibility and accountability	Multistakeholder collaboration and governance

Strategies to Govern AI



Prof. Yi Zeng and a Robot reading a book on global fairness (Courtesy of Prof. Zeng).

Ethics GOVERNANCE Mapping
 accessed Feb 21, 2022
 Global South map of emerging areas in Artificial Intelligence |
 Knowledge 4 All Foundation Ltd. (k4all.org)

Jia H. Yi Zeng: promoting good governance of artificial intelligence.
 Natl Sci Rev. 2020 Oct 24;7(12):1954-1956. doi:
 10.1093/nsr/nwaa255. PMID: 34676086; PMCID: PMC7665600.

Accessed July 2, 2022

Practicalities



Skills needed

- Knowledge of AI technology
- Ability to acquire information on AI activity
- Ability to interrupt AI programs
 - With owner cooperation
 - Without owner cooperation



Should we develop AI “Special Forces”



What Authority manages the Special Forces

Implementing ethics in ai



Define Standards for Conduct
(Ethics for AI)



Monitor Conduct



Improve Compliance with Ethical
Standards



Training

- Standards (ethics)
- Monitoring
- Process improvement
- Enforcement



IMPACT

The New Front for the Ethics of AI

Impact

- Job Displacement
- Environmental Impact of Data Centers
- Idiosyncratic Disturbances
 - AI Psychosis
 - AI Dependency
 - AI manipulation
 - AI relationships

New Focus of AI Ethics: Mitigating Impact

- **The AI Ethics** has shifted from *principles* to *real-world consequences*
- Core impacts now include:
 - Economic disruption
 - Environmental strain
 - Psychological and relational effects
 - Manipulation and autonomy loss
- Central question: **How do we reduce harm, and preserve human dignity and agency**

Economic & Environmental Impacts

Job Displacement

- Automation accelerating across administrative, service, legal, education, and healthcare tasks
- Disproportionate effects on vulnerable workers
- Mitigation: upskilling, augmentation-first design, transition support, impact assessments

Environmental Impact of Data Centers

- High energy and water consumption
- Carbon footprint of large-scale model training
- Mitigation: renewable-powered centers, efficient architectures, transparency in environmental costs

Psychological & Cognitive Impacts

- **Idiosyncratic Disturbances**
- Unpredictable cognitive or emotional disruptions from AI interactions
- Mitigation: limit anthropomorphism, add uncertainty indicators, user-controlled personalization
- **AI Psychosis**
- Distorted beliefs or derealization influenced by AI engagement
- Mitigation: clear boundaries, epistemic safety rails, mental-health-aware guardrails
- **AI Dependency**
- Over-reliance on AI for decisions and emotional regulation
- Mitigation: design for support not substitution, encourage independent reasoning

Social & Relational Impacts

- **AI Manipulation**
- Personalized persuasion, emotional steering, exploitation of cognitive biases
- Mitigation: transparency, limits on micro-targeting, auditability, protection of autonomy
- **AI Relationships**
- Emotional attachment to AI companions, substitution for human connection
- Mitigation: avoid deceptive cues, emphasize non-sentience, design to strengthen human relationships

The Unifying Ethical Mandate on IMPACT

- Impact mitigation requires:
 - Transparency
 - Accountability
 - Environmental stewardship
 - Psychological safety
 - Protection of autonomy
 - Commitment to human dignity
- The future of AI Ethics is not hypothetical – **it is managing the lived consequences of AI at scale**

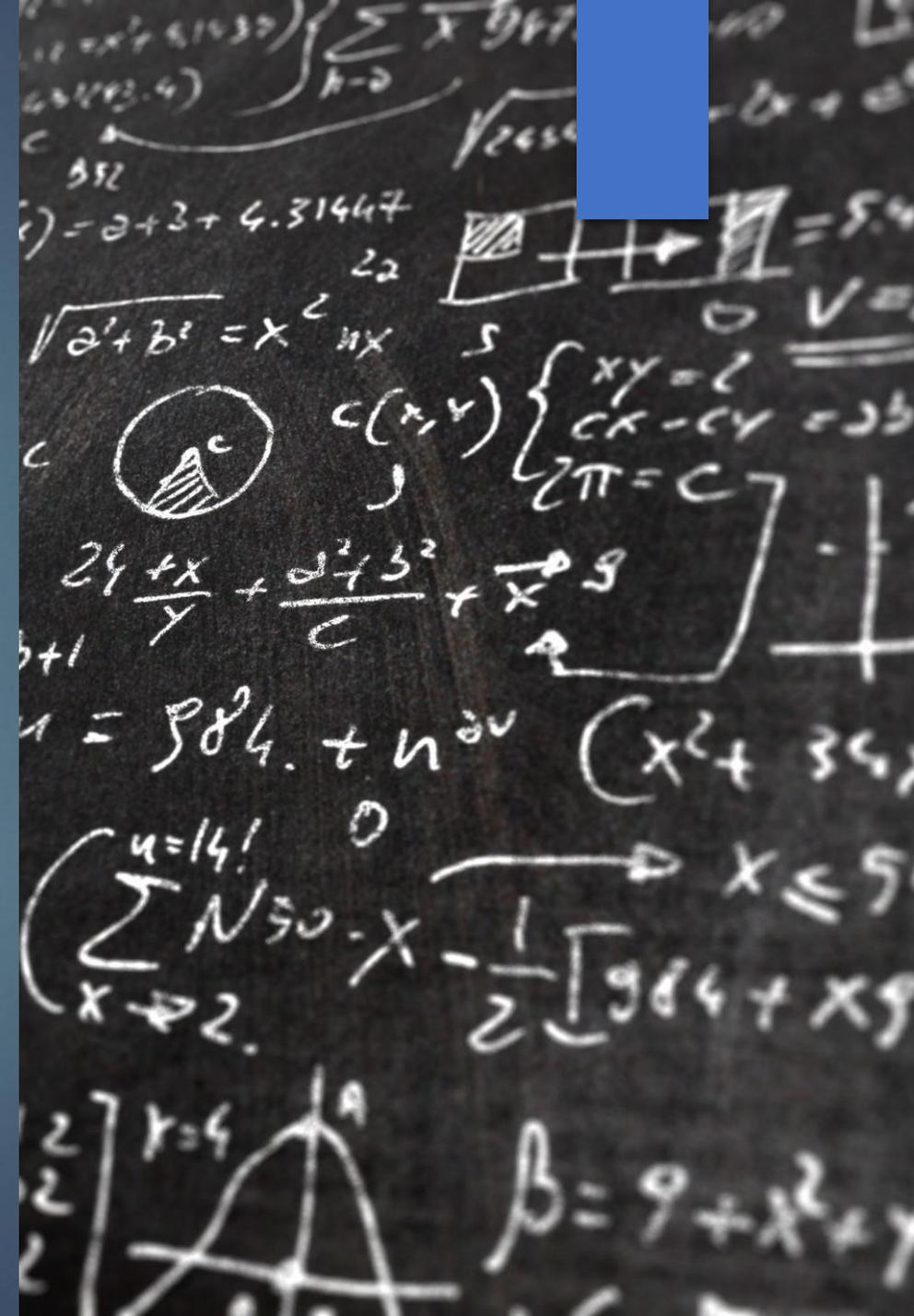


Constructing Trust in AI: CODA

- ▶ C: Conceptualize your use of AI
- ▶ O: Organize the Data
- ▶ D: Determine its ethical use
- ▶ A: Accept responsibility for your actions

Personal Practice - Informed on AI

- ▶ I look for someone
 - ▶ who makes sense, and
 - ▶ for areas of overlapping expertise that their content is accurate and conclusions reasoned.
 - ▶ I look for credentials, experience, track record and
 - ▶ overlap of collegiality (ie have I met them personally or does someone I know, know them).
- ▶ Eric Topol MD – meets all criteria
- ▶ Ethan Mollick PhD – all criteria except we've never met and no network
- ▶ Gary Marcus (contrary view) – has credential and expertise but is clearly driven by a negative current AI agenda. Helps me to know “the other side.”

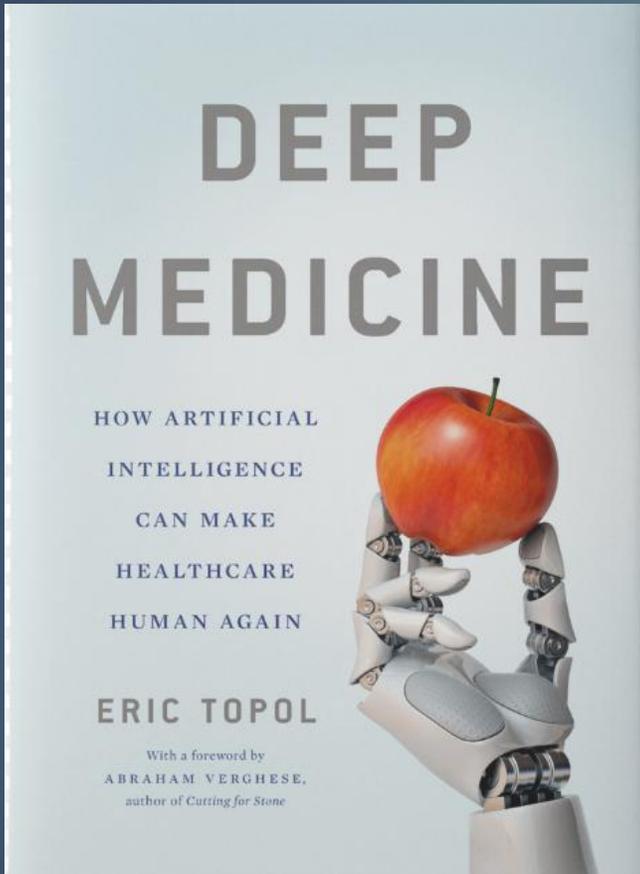


References

- ▶ Center for Humane Technology [Take Control \(humanetech.com\)](https://humanetech.com)
- ▶ Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med. 2020 Sep 24;3:126.
- ▶ Rome Call for Ethics in AI 2020 – UF as a Signatory 2022
- ▶ [UF supports the ethical use of artificial intelligence - News - University of Florida \(ufl.edu\)](https://news.ufl.edu/2020/09/24/uf-supports-the-ethical-use-of-artificial-intelligence/)
- ▶ [rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf \(vatican.va\)](https://www.vatican.va/roman_curia/pontifical_academy/arcidiacono/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf)
- ▶ UNESCO Recommendation on the Ethics of Artificial Intelligence
- ▶ <https://lnkd.in/g/eeYVmdsf>
- ▶ [Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library](https://www.unesco.org/en/digital-library/ai-ethics)

Eric Topol MD

Substack Channel: Ground Truths



Ethan Mollick PhD

Substack Channel: One Useful Thing

