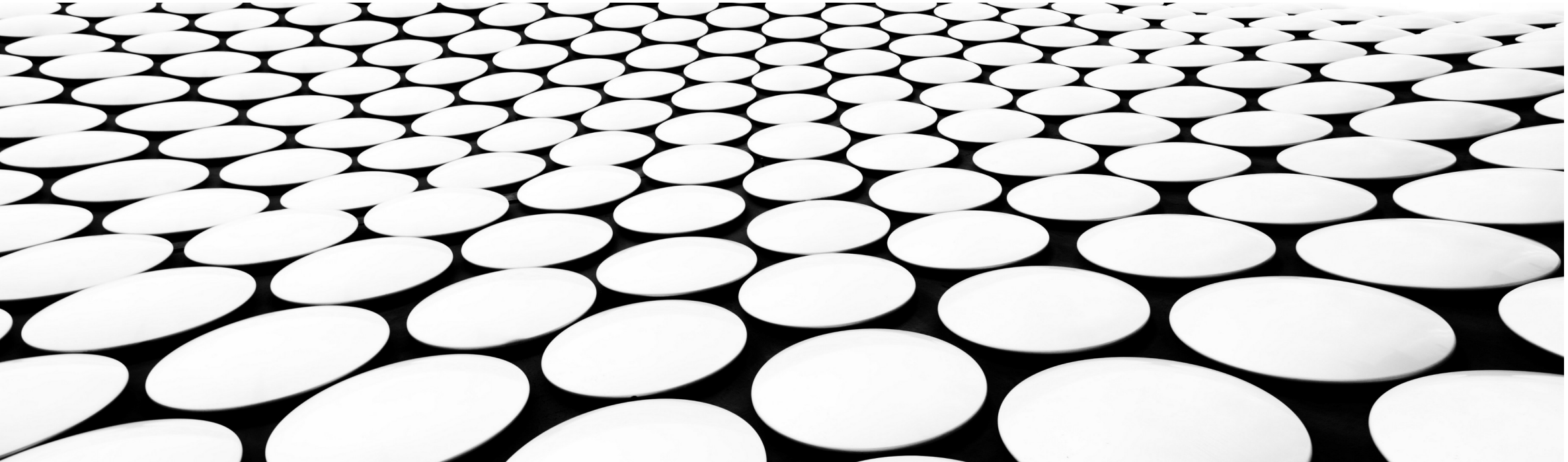

ETHICS IN AI HEALTH

YING DING, SCHOOL OF INFORMATION, DELL MEDICAL SCHOOL, UT AUSTIN



MINDSET SHIFT: AUGMENT VS. AUTOMATE/REPLACE



Augment

- Bike
- Car



?? (Happy to be replaced) Automate/Replace

- Washing machine
- Dishwasher
- Driverless Car
- ChatGPT



Erik Brynjolfsson

Director

Stanford Digital Economy Lab

HUMAN PRIVILEGED SKILLS

Narrow AI vs. General AI

- Thinking
- Creating
- Judging
- Deciding
- Diagnosing
- Inventing



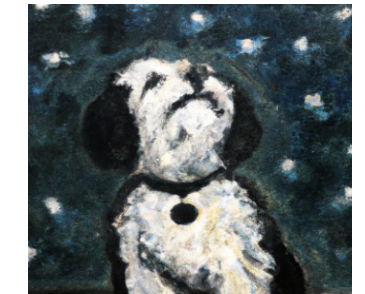
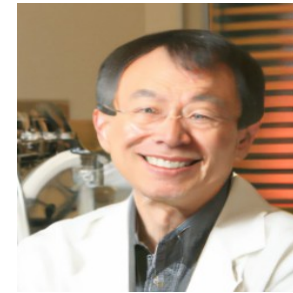
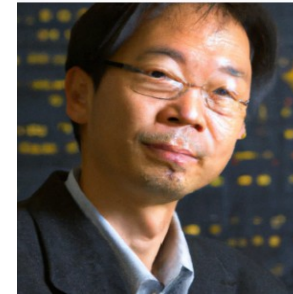
- Aggregating info to make decision: Leap vs. Step



AI ethics and trust are important considerations in the design and use of artificial intelligence systems. AI ethics refers to the ethical principles and values that should guide the development and deployment of AI systems, while trust in AI refers to the extent to which people have confidence in the reliability and fairness of these systems.

There are a number of ethical concerns that must be considered when developing and using AI, including issues related to bias, transparency, accountability, and fairness. AI systems can perpetuate and amplify existing biases if they are trained on biased data, and they can also be used to make decisions that have significant consequences for people's lives, such as hiring decisions or parole recommendations. As a result, it is important that AI systems are designed and used in a way that is transparent, accountable, and fair, and that takes into account the potential impacts on different groups of people.

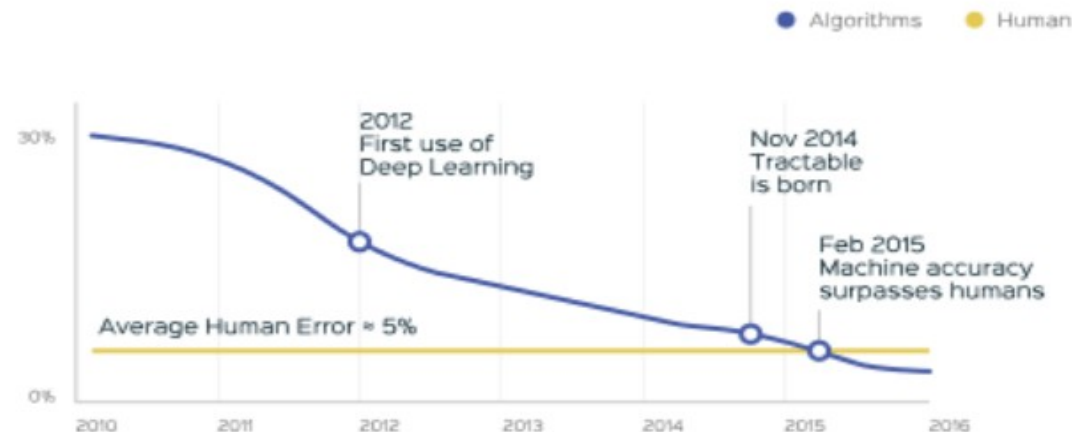
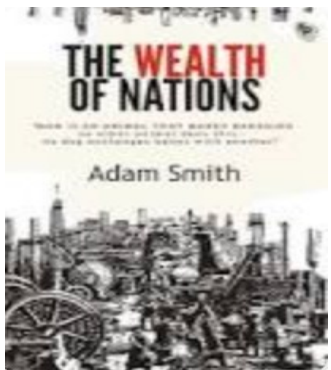
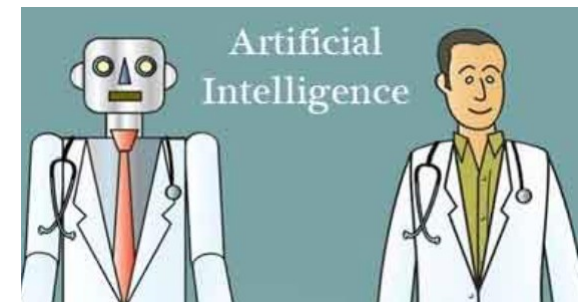
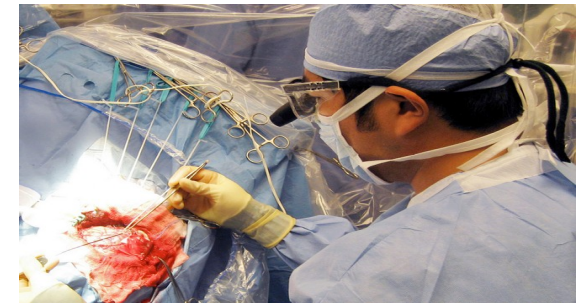
Building trust in AI is also important, as people are more likely to use and rely on AI systems if they have confidence in their reliability and fairness. This requires ensuring that AI systems are transparent in their decision-making processes, and that they are tested and validated to ensure that they are accurate and unbiased. It also requires ensuring that AI systems are accountable for their actions and decisions, and that there are mechanisms in place to address any negative impacts they may have.



AI DOCTOR VS. HUMAN DOCTOR



An AI doctor is a type of artificial intelligence system that is designed to assist with medical tasks, such as diagnosing patients or recommending treatment options. These systems can use a variety of techniques, including machine learning, natural language processing, and image recognition, to analyze medical data and provide insights or recommendations to healthcare professionals. Some AI doctors are designed to work independently, while others are intended to assist human doctors in their work. There is ongoing research and development in this area, and it is expected that AI doctors will play an increasingly important role in healthcare in the future. However, it is important to note that AI systems should be used as a supplement to, rather than a replacement for, human medical professionals.



IN GOD WE TRUST! IN AI WE TRUST? WHOM TO BELIEVE IN?

- AI does not have the capacity to be trusted because it does not possess emotive states or be held responsible for its actions ⇒ requirements of the affective and normative accounts of trust.
- AI does meet all the requirements of the rational account of trust ⇒ a form of reliance.
 - Rational trust is that the trustor is making a logical choice
- German philosopher Martin Heidegger said that the human-technology relationship oscillates between suspicion and reliance.
- **Being trustworthy**



If you were injured in a car accident caused by a defect or repair mistake in Texas, you may be able to pursue compensation from the manufacturer or auto repair shop through an insurance claim or lawsuit.

4 PILLARS OF MEDICAL ETHICS

THE PILLAR OF
BENEFACTENCE



THE PILLAR OF
NON-MALEFACTENCE



THE PILLAR OF
AUTONOMY



THE PILLAR OF
JUSTICE



The four pillars of medical ethics are:

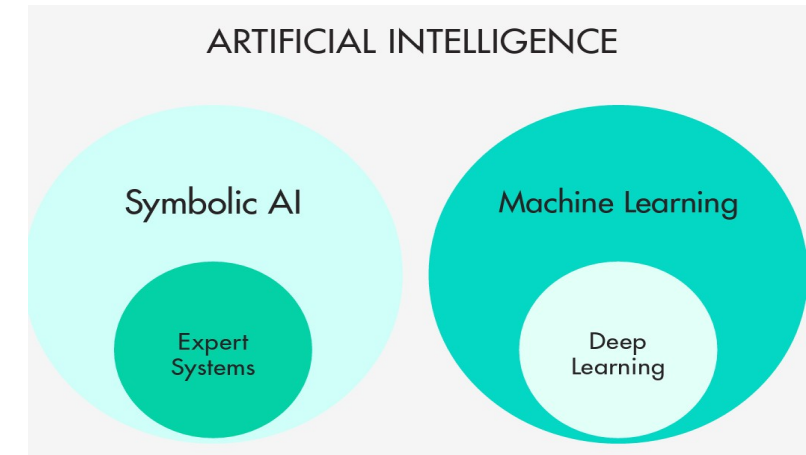
- Beneficence (doing good)
- Non-maleficence (to do no harm)
- Autonomy (giving the patient the freedom to choose)
- Justice (ensuring fairness)

<https://link.springer.com/article/10.1007/s11948-020-00228-y>
<https://link.springer.com/article/10.1007/s43681-022-00200-5>

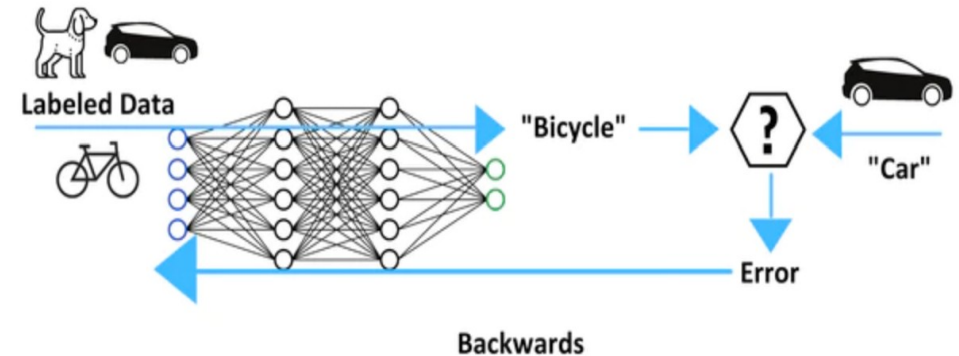
AI IN HEALTHCARE

- AI can diagnose skin cancer better than board-certified dermatologists
- Chatbots are used to dispense medical advice in lieu of nurses
- Machine learning methods for sepsis prediction
- Medical imaging diagnosis
- Surgery robots
- Genomic biomarker identification for cancer patients
- Stroke diagnosis
- Medical notes summary
- Drug discovery
- Clinical trials
- Hospital management
- Personalized care management - health insurance company

Whether AI can use data safely, securely, and appropriately and to be deployed ethically and effectively



Deep Learning Training



- Data Bias
- Model Transparent
- Result Interpretation - Contextualization/Personalization

AI IN HEALTHCARE

Use Cases	Examples	
	Current	Future
1. Process optimisation <i>Using AI to optimise 'back-end' processes in healthcare, such as procurement, logistics, and staff scheduling</i>	<ul style="list-style-type: none"> Rota/staff schedule management, e.g. Hong Kong Health Authority Emergency services dispatch management e.g. Corti 	<ul style="list-style-type: none"> Data-driven optimisation of logistics, procurement Automated analysis/ completion of medical notes and other documentation Patient experience analysis e.g. Alder Hey
2. Preclinical research <i>Using AI in preclinical applications such as drug discovery and genomic science</i>	<ul style="list-style-type: none"> Candidate molecule screening, e.g. BenevolentAI, AtomNet Repurposing drugs, e.g. Teva Pharmaceuticals Predicting potential side effects, e.g. Cloud Pharmaceuticals 	<ul style="list-style-type: none"> Determining targets for gene editing, e.g. CRISPR

3. Clinical pathways

Fitting AI into existing and new clinical workflows, such as in diagnostics and prognostication

- Analysis of optical coherence tomography (OCT) images, e.g. DeepMind-Moorfields collaboration
- Analysis of radiological imaging, e.g. Viz.ai
- Analysis of clinical conversations e.g. Corti
- Prognostication e.g. prediction of all-cause mortality [Stanford, KenSci], prediction of cardiovascular risk [University of Nottingham]
- Radiologists' assistants, e.g. suggesting best imaging modality in particular clinical situation, improved image acquisition processes leading to radiation dose reduction
- Management decision-support for healthcare practitioners, suggesting best treatment for particular patient
- Automated transcription of clinical interactions
- Automated completion and submission of investigation requests/referrals

4. Patient-facing applications

Using AI to interact directly with patients and other service users, including in the delivery of therapies or the provision of information

- Chatbots, e.g. Oli [Alder Hey], AVA [Arthritis Research UK], Lark Weight Loss Coach
- Autonomous (closed-loop) insulin pumps
- Personalised health advice and interventions, e.g. CareSkore, Viome, DayTwo
- Smart homes and wearables
- Robot carers
- Robot surgeons

5. Population-level applications

Using AI to gain insights into population health, such as identifying epidemics and monitoring disease spread.

- Prediction of infectious disease outbreaks, e.g. Dengue fever app in Malaysia
- Better targeting of public health spending and other interventions, e.g. University of Southern California tool
- Better understanding of risk-factors for non-communicable disease, e.g. childhood obesity [Indiana University tool]

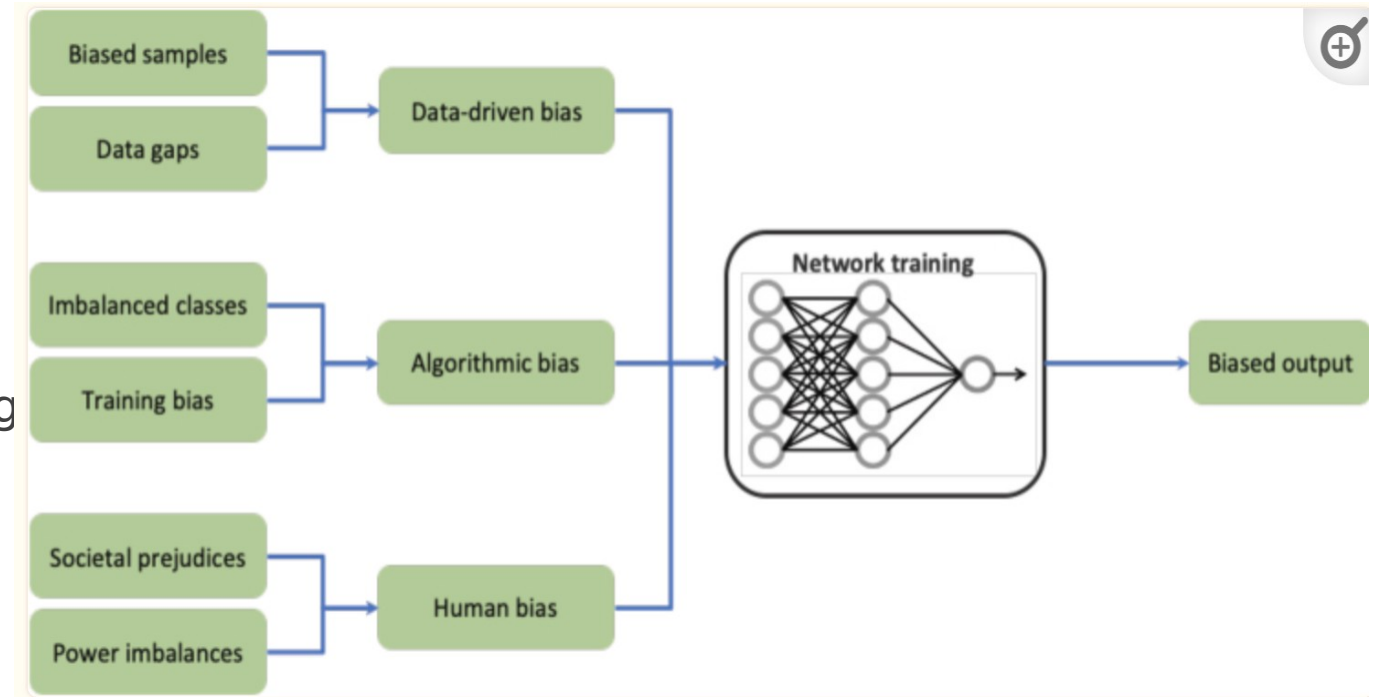
A summary of the epistemic, normative and overarching ethical concerns related to algorithmic use in healthcare based on [Mittelstadt et al. \(2016\)](#) from ([Morley and Floridi, 2020b](#)).

	Ethical Concern	Explanation	Medical Example
Epistemic concerns	Inconclusive Evidence	Algorithmic outcomes (e.g. classification) are probabilistic and not infallible. They are rarely sufficient to posit the existence of a causal relationship.	<i>EKG readers in smartwatches may 'diagnose' a patient as suffering from arrhythmia when it may be due to a fault with the watch not being able to accurately read that user's heartbeat (for example due to the colour of their skin) or the 'norm' is inappropriately calibrated for that individual (Hailu, 2019).</i>
	Inscrutable Evidence	Recipients of an algorithmic decision very rarely have full oversight of the data used to train or test an algorithm or the data points used to reach a specific decision.	<i>A clinical decision support system deployed in a hospital may make a treatment recommendation, but it may not be clear on what basis it has made that 'decision' raising the risk that it has used data that are inappropriate for the individual in question or that there is a bug in the system leading to issues with over or under prescribing (Wachter, 2015).</i>
	Misguided Evidence	Algorithmic outcomes can only be as reliable (but also as neutral) as the data they are based on.	<i>Watson for Oncology is in widespread use in China for 'diagnosis' via image recognition but has primarily been trained on a Western data set leading to issues with concordance and poorer results for Chinese patients than their Western counterparts (Liu et al., 2018).</i>
Normative Concerns	Unfair outcomes	An action can be found to having more of an impact (positive or negative) on one group of people	<i>An algorithm 'learns' to prioritise patients it predicts to have better outcomes for a particular disease. This turns out to have a discriminatory effect on people within the Black and minority ethnic communities (Garattini et al., 2019).</i>
	Transformative effects	Algorithmic activities, like profiling, re-conceptualise reality in unexpected ways.	<i>An individual using personal health app has limited oversight over what passive data it is collecting and how that is being transformed into a recommendation to improve, limiting their ability to challenge any recommendations made and a loss of personal autonomy and data privacy (Kleinpeter, 2017).</i>
Overarching	Traceability	Harm caused by algorithmic activity is hard to debug (to detect the harm and find its cause), and it is hard to identify who should be held responsible for the harm caused.	<i>If a decision made by clinical decision support software leads to a negative outcome for the individual, it is unclear who to assign the responsibility and/ or liability to and therefore to prevent it from happening again (Racine et al., 2019).</i>

Morley, J., Machado, C. C. V., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2022). The ethics of AI in healthcare: A mapping review. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3830408

DATA BIAS AND TRUST: TRANSLATIONAL

- Bias in patient population distribution
- Rare diseases
- Lack of high-quality annotated data
- Data privacy and sharing
- Cross-modality of data: table, text, image, wave
- Missing data
- Genomic data (cannot be de-identified)



- **Not all data are created equally**
- **Data Governance**
- **Data Standard**

African-Americans comprise less than 5% of clinical trial participants and Hispanics just 1%, in spite of the fact that they represent 12% and 16% of the total US population respectively.

THE FALL OF IBM'S WATSON ONCOLOGY

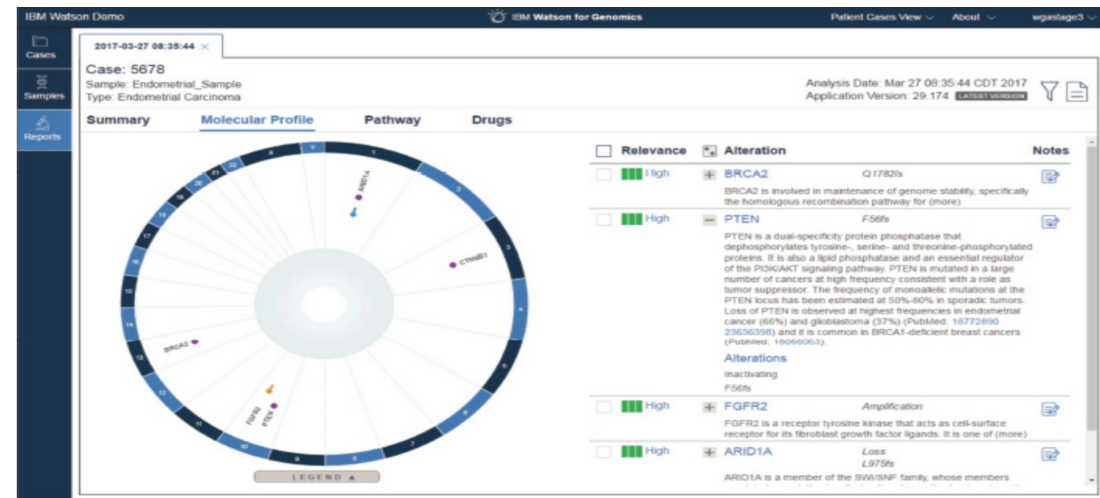
Product Life Cycle Dates

Program Number	VRM	Announced	Available	Marketing Withdrawn	Service Discontinued
5725-W51	00.00.00	2015-06-23	2015-06-23	2020-12-31	2022-07-01

↑ Back to top

Program number

- IBM Watson for Oncology SaaS (5725-W51)



In 2012, IBM closed a deal with Memorial Sloan Kettering, one of the preeminent cancer centers in the country, to train an AI to make treatment recommendations.

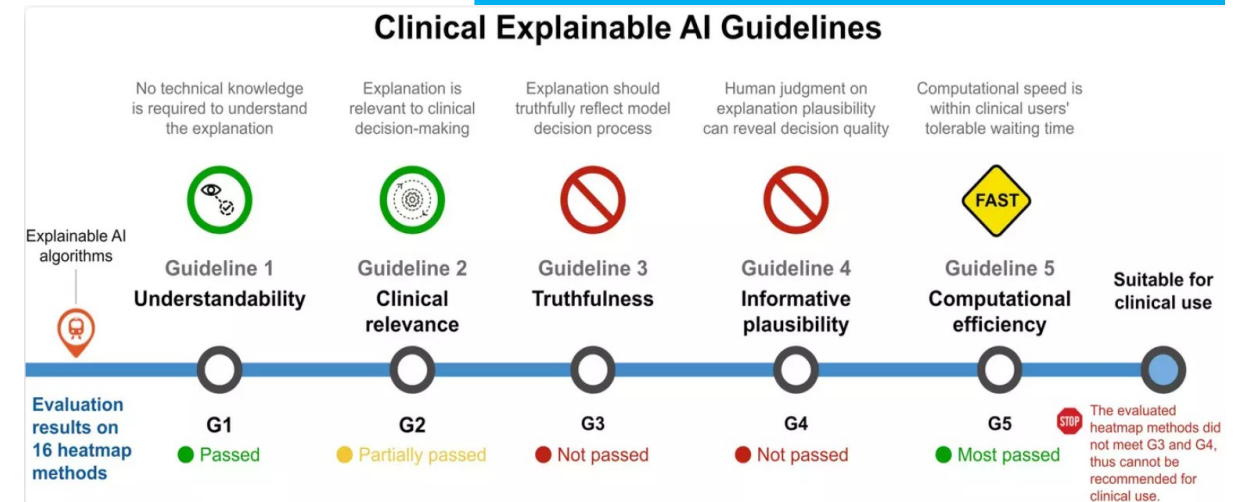
Truven, Phytel, Explorys and merge. Truven had the biggest insurance database in the nation with 300 million covered lives, Explorys provided a clinical data set of actual electronic health records kept by health systems representing about 50 million or so patients, Phytel added on top of that, and Merge had a huge imaging database. They had all this data and the idea was: Expose Watson to that, and it finds patterns that physicians and anyone else can't possibly find when looking at that data, given all the variables in it.

ALGORITHM TRANSPARENCY: EXPLAINABILITY

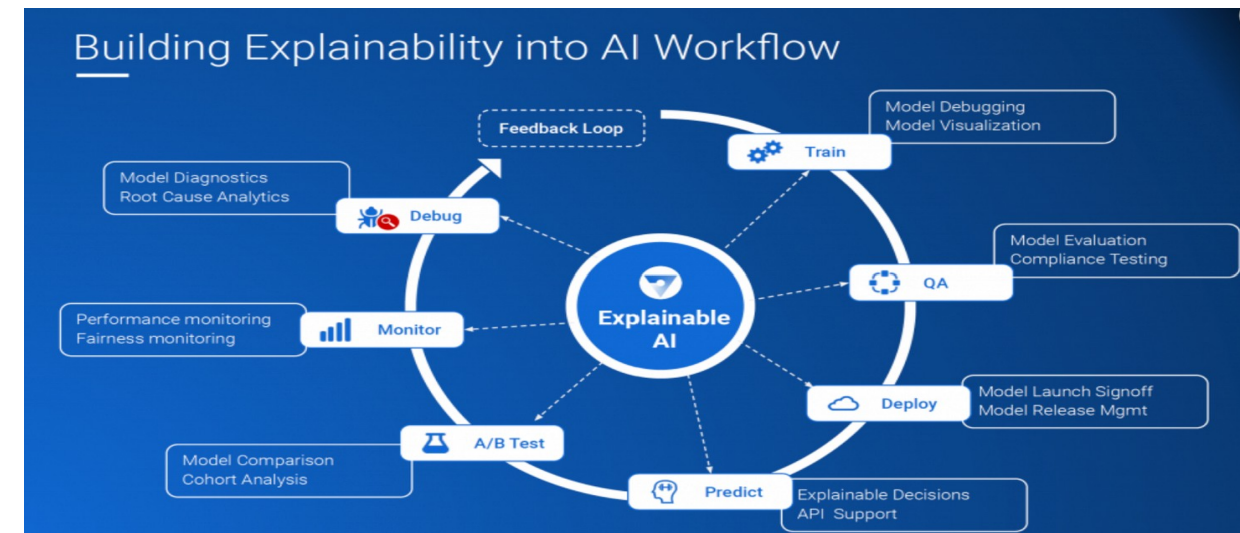
Is this human 'black box' very different from the algorithmic black box?

- The right to explanation
- Different users and different situations will require different things in terms of explainability.
- Counterfactual explanations: what if, why not,
- Decisions are based on a combination of information and preferences, user preferences are ignored.
- Human error vs. algorithmic error: we hold algorithms to a higher standard than we hold humans
- Court: judge decisions/actions as negligent or malicious through court discussion and human explanation, but machine cannot join court discussion
- Interaction: itself is a process of explanation

<https://www.uxai.design/design-strategy>



https://weina.me/clinical_xai_guideline



<https://www.fiddler.ai/blog/explainable-ai-goes-mainstream-but-who-should-be-explaining>

CONTEXTUALIZING AI RESULTS: ETHICS OF CAREBOTS

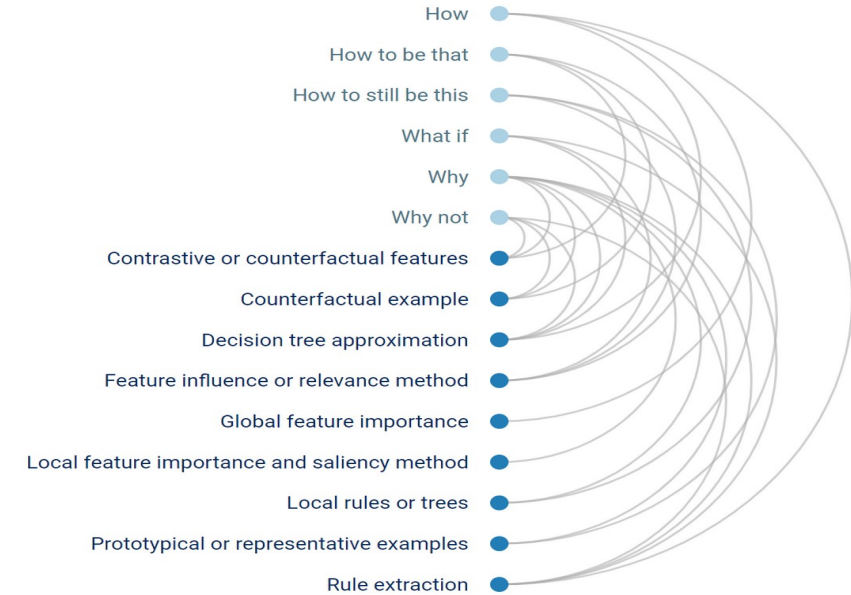
- **Balancing:** Overmedicalising healthy people, causing anxiety – prediabetes
- **Localization or personalization:** from generalization (what AI is good at) to degeneralization (what AI is not good at)
- Ethics of Carebots:
 - infringements of privacy
 - Deception for early with dementia
 - Overreliance and attachment
 - maintenance of trust
 - Deontology: issues of autonomy, dignity, deception, and social isolation
 - Health disparity: affordability of carebots
 - Human Dignity



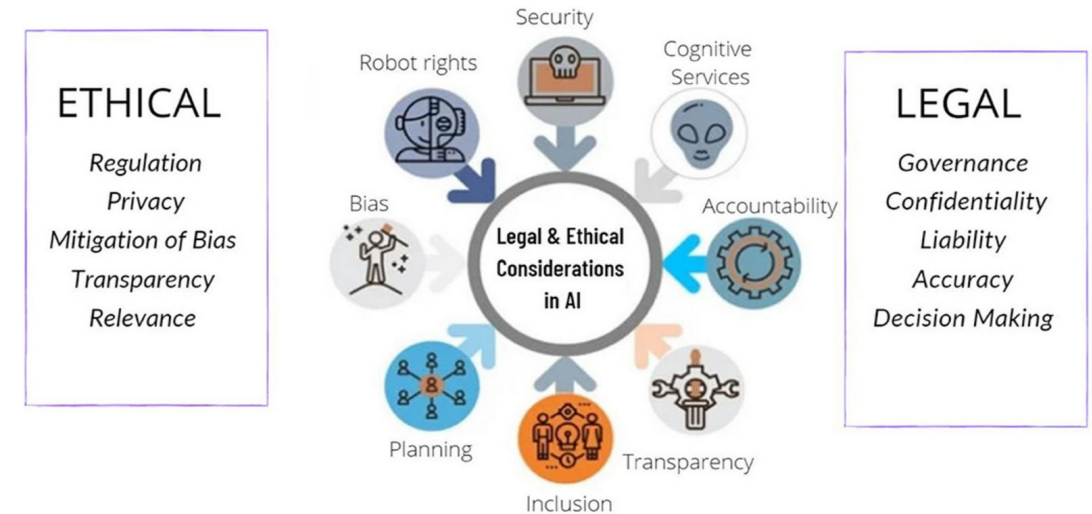
Who is to be held accountable if a care receiver takes his own life under the watch of a care robot which is an autonomous agent.

DESIGN: A TRUST BRIDGE BETWEEN AI, DOCTOR, AND PATIENT

- Including end-users (e.g., clinicians) in the design of AI in healthcare
- Being explainable and transparent on AI technologies
- The involvement of end user in the design can reinforce transparent development process.
- Ethical Design
 - top-down: embedding ethical principles into AI algorithm
 - bottom up: AI algorithm adapts and learns an external set of principles based on the observations of humans
 - technomoral care: skillful, attentive, responsible and emotionally responsive disposition to personally meet the needs of others who share our technosocial environment



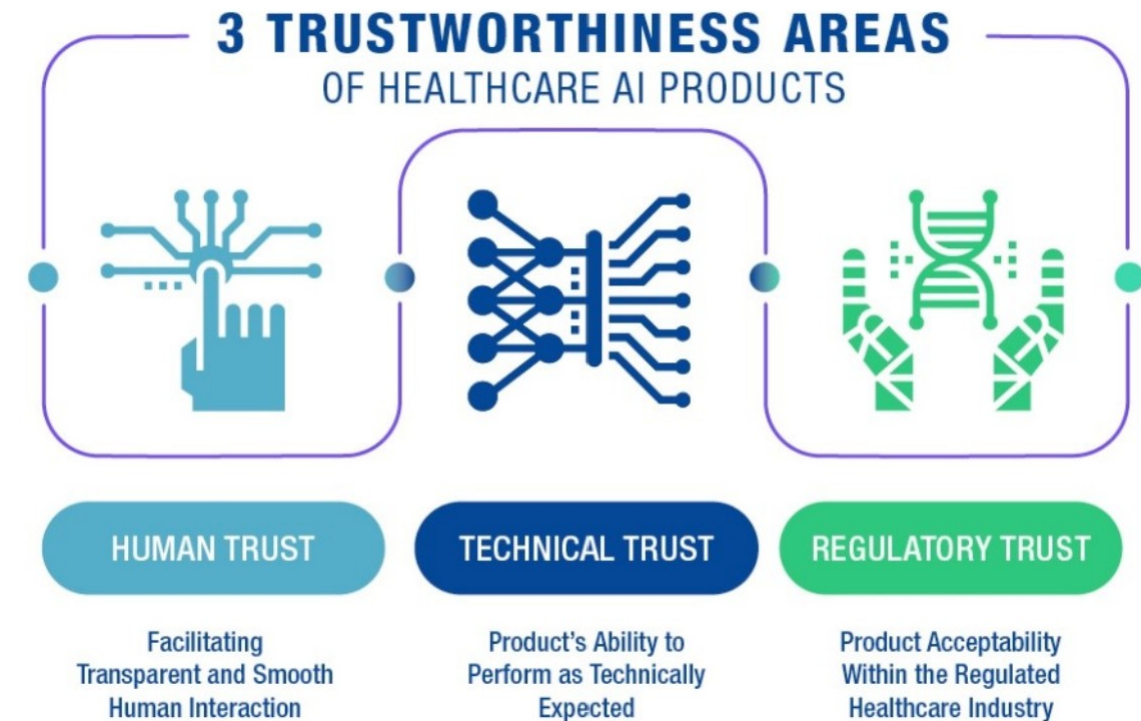
<https://www.uxai.design/design-strategy>



<https://www.frontiersin.org/articles/10.3389/fsurg.2022.862322/full>

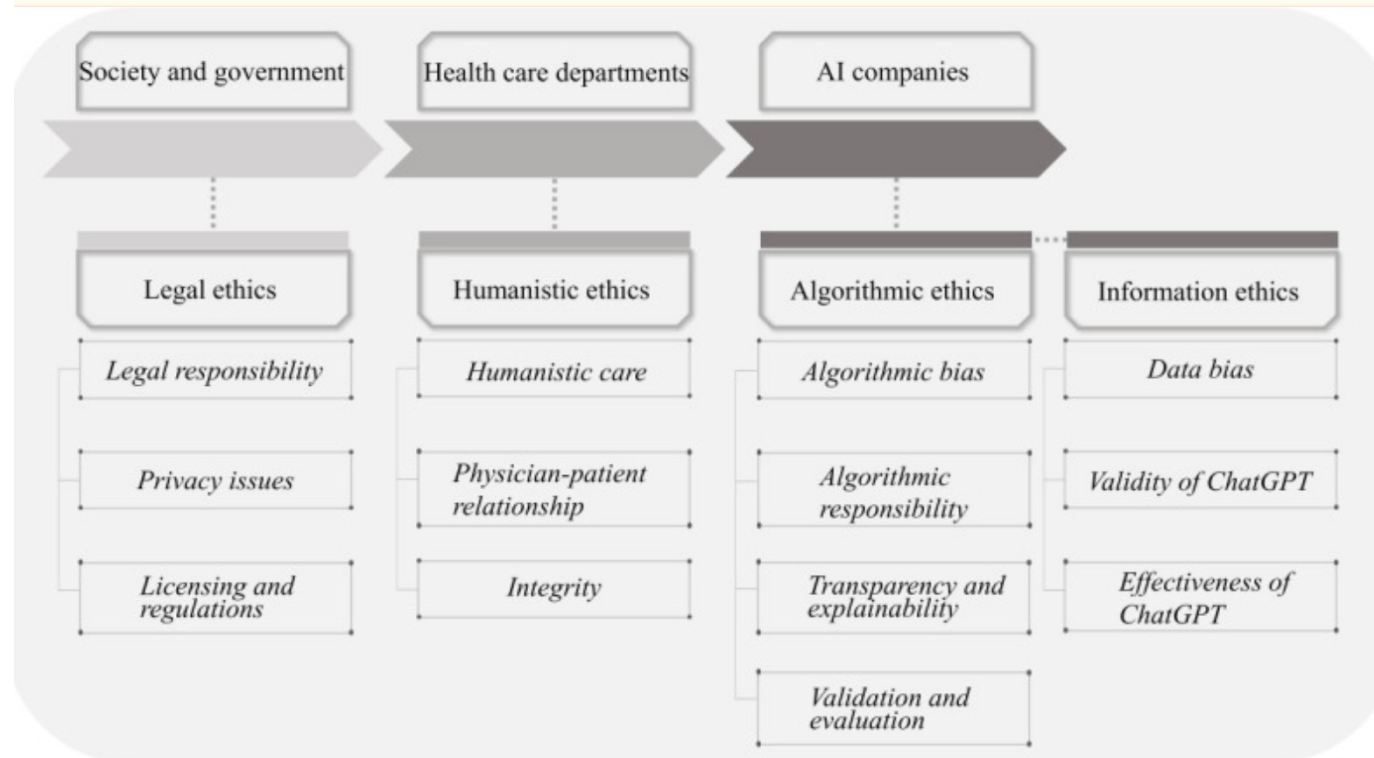
DEMISTIFYING AI

- **Trust** is a relational and normative concept.
 - **Trust** implies some uncertainty or risks
 - **Human and Technology Trust:** iterative process of testing and evaluating
 - **Trust** is a dynamic process
-
- Communication
 - Education
 - Governance: developing policy to address ethical, social, and legal issue
 - Regulation: continuous certification for AI
 - Monitoring: learning, adapting, and mitigating bias



CHATGPT AND ETHICAL AI IN HEALTHCARE

ChatGPT may need to hold a valid license and comply with specific regulations governing their practice.



Humanistic ethics should guide the use of ChatGPT in health care, emphasizing the importance of a person-centered approach.

Mitigate the potential “automation bias” that can result from overreliance on algorithms

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10457697/#:~:text=ChatGPT%20presents%20potential%20ethical%20challenges,privacy%20due%20to%20data%20collection.>

HOW CHATGPT THINKS ABOUT THIS



you are an AI ethics expert. Can you talk about few important perspectives about ethical issues of using ChatGPT for healthcare



Certainly. The use of AI models, like ChatGPT, in healthcare presents a myriad of ethical considerations. Here are some important perspectives on the ethical issues:

1. Data Privacy and Confidentiality:

- Healthcare discussions often involve sharing personal health information (PHI). Ensuring the privacy and confidentiality of this data is paramount.
- Even if the data isn't stored, the mere processing of sensitive health information by commercial entities could pose risks and concerns.

2. Reliability and Accuracy:

- Misinformation or misinterpretation by an AI model can have direct health consequences.
- The stakes in healthcare are incredibly high; incorrect information can lead to misdiagnosis, inappropriate treatments, or delay in care.

3. Equity and Bias:

- AI models are trained on vast amounts of data. If the data is biased, the model's responses will reflect those biases.
- In healthcare, this could mean that certain populations receive incorrect or less effective care advice based on biased data.

Regenerate

4. Transparency:

- Understanding how an AI model reaches its conclusions is crucial, especially in healthcare.
- Without transparency, it's challenging to trust, verify, or validate the model's recommendations.

5. Accountability and Liability:

- If a patient acts on the advice of an AI and experiences harm, who is held accountable? The developers? The healthcare provider that deployed the AI?
- The boundaries of liability need clear definition when using AI in health contexts.

6. Autonomy and Informed Consent:

- Patients have the right to understand the tools involved in their care.
- They should be informed that an AI is involved, understand its capabilities and limitations, and give consent to its use.

7. Continuity of Care:

- Relying on AI for advice might disrupt the patient-provider relationship. Continuous care, understanding a patient's history, and building trust are crucial in healthcare.
- AI can't replace the human touch, understanding, and rapport necessary in many healthcare situations.

8. Over-reliance:

- If healthcare providers or patients come to depend too heavily on AI tools like ChatGPT, they might ignore their intuition, training, or knowledge.
- AI should complement, not replace, human judgment.

Regenerate

HOW CHATGPT THINKS ABOUT ETHICAL AI IN HEALTHCARE

9. **Accessibility:**

- While AI can make healthcare advice more accessible, there's also a risk of widening health disparities if only certain populations have access to AI-enhanced care.

10. **The "Black Box" Dilemma:**

- Deep learning models, including ChatGPT, are often viewed as "black boxes" because their internal workings and decision-making processes are not easily explainable.
- In healthcare, where understanding the rationale behind decisions can be critical, this poses an ethical challenge.

In conclusion, while AI models like ChatGPT offer immense potential in revolutionizing healthcare with instant information and assistance, they also bring forth significant ethical challenges. Addressing these concerns requires a concerted effort from technologists, ethicists, healthcare professionals, and policymakers.

LONG WAY TO GO

- In 1994, Betsy Lehman, a reporter of Boston Globe and mother of two young girls, was battling advanced stage of breast cancer. She was killed by the four times the intended dose of a powerful chemotherapy drug at Dana-Farber Hospital (the world premier cancer center). – Betsy Lehman Center



<https://betsylehmancenterma.gov/>