生物医学工程学院
School of Biomedical Engineering

上海科技大学
ShanghaiTech University

Medical Data Privacy and Ethics in the Age of Artificial Intelligence

Lecture 2: Overview (AI Ethics)

Zhiyu Wan, PhD (wanzhy@shanghaitech.edu.cn)

Assistant Professor of Biomedical Engineering

ShanghaiTech University

September 19, 2025

Goals of this lecture

- After this lecture, students will learn:
  - Principles of Ethical AI

XPRIZE Video: AI for Good - AI and Medicine

- Artificial intelligence will change lives in many ways. Already, AI solutions are being deployed and having significant impact in healthcare. Daniel Kraft, MD, Chair for Medicine, Singularity University, shares his expert views on how significant this technology will be in finding the right diagnosis and therapies and shifting the ‘practice’ of medicine to the real ‘science’ of medicine. (May 3, 2017)

XPRIZE Video: AI for Good - Ethics in AI

- Address AI from an ethics, safety, moral and privacy rights perspective and the need for a guiding ethical framework and code of conduct to create a foundation for the design, production and use of AI. (May 27, 2017)

Definitions

- **Ethics**: The systematic set of principles or guidelines that govern conduct within a particular context, often derived from philosophical theories and professional standards. It deals with questions of what is right and wrong, and what individuals ought to do in various situations.
- **Values**: Individual or collective beliefs about what is important, desirable, and worthwhile. They reflect personal or cultural priorities and inform decisions and behaviors.
- **Morality**: The principles or rules of behavior that individuals or societies believe are right or wrong. It is more focused on practical, everyday conduct and often carries a strong emotional and social component.

Etymology of Ethics

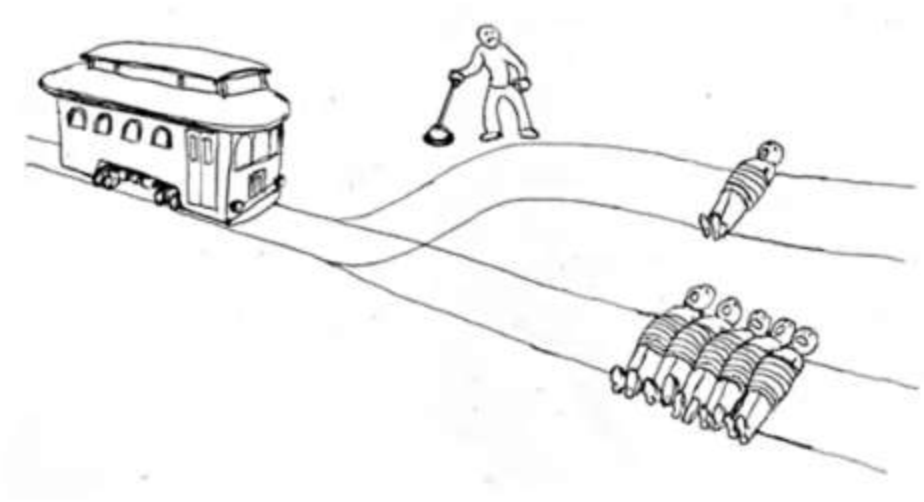
- Etymologically, the English word “ethics” (ethica in Latin) can be traced back to the ancient Greek noun, (ethos), which denotes a “habit” or “custom”.
- Ethics is a practice discipline that refers to human action with the purpose of being morally good.
- In Chinese, “伦” means “Order” which represents relationships; “理” means “Rule” which represents principles.
- “Morality” is more subjective. “Ethics” is more objective.

Introduction to Ethics

- **Definition:** Ethics is the branch of philosophy that deals with questions about what is morally right and wrong, fair and unfair, good and bad.
- **Purpose**: To guide human behavior, ensuring individuals and organizations act in a morally responsible way.
- **Healthcare Ethics**: Ensures that professionals make decisions in the best interests of patients (e.g., informed consent, confidentiality).
- **Technology Ethics**: Ethical concerns regarding data privacy, AI development, and the impact of technology on society.

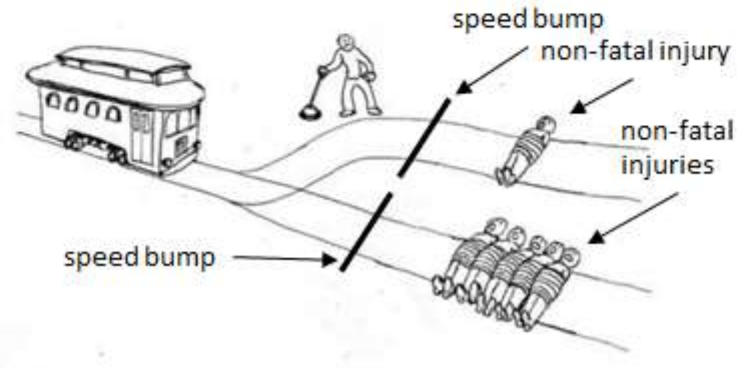
What would you do?

- **Trolly Problem**
  - Imagine a runaway trolley speeding down the tracks toward five people who are tied up and unable to move. You are standing beside a lever that can divert the trolley onto another track. But on that side track, there is one person tied up. You now face a difficult choice: do nothing and allow the trolley to kill the five people, or pull the lever and sacrifice one person to save the five.



What would you do?

- Variation 1: The Non-fatal Switch

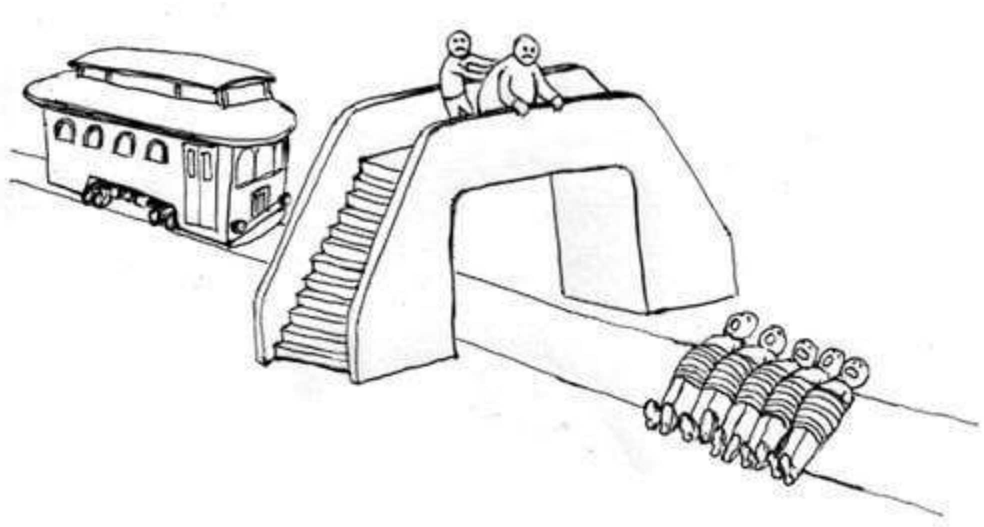


The non-fatal switch

https://themindcollection.com/trolley-problem-meme-variations/

What would you do?

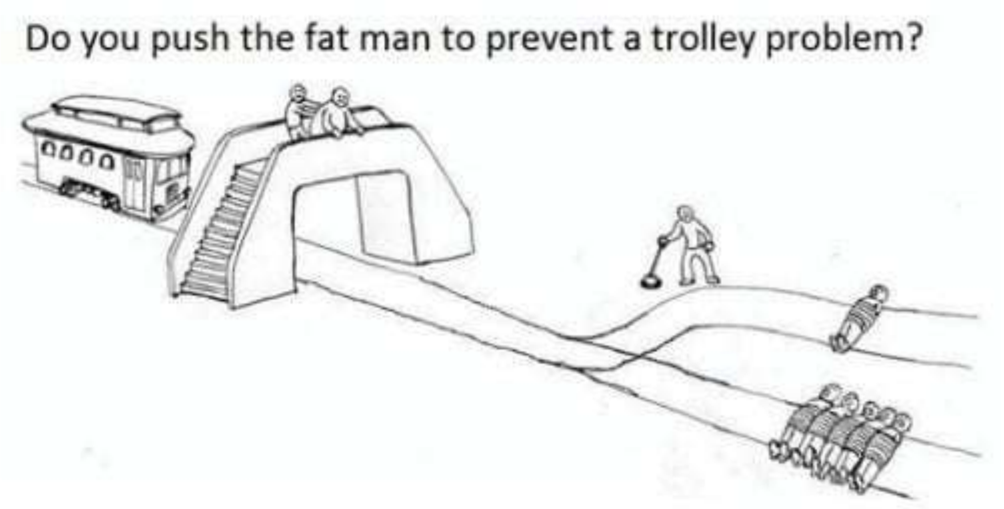
- Variation 2: The Fat Man Problem



https://themindcollection.com/trolley-problem-meme-variations/

What would you do?

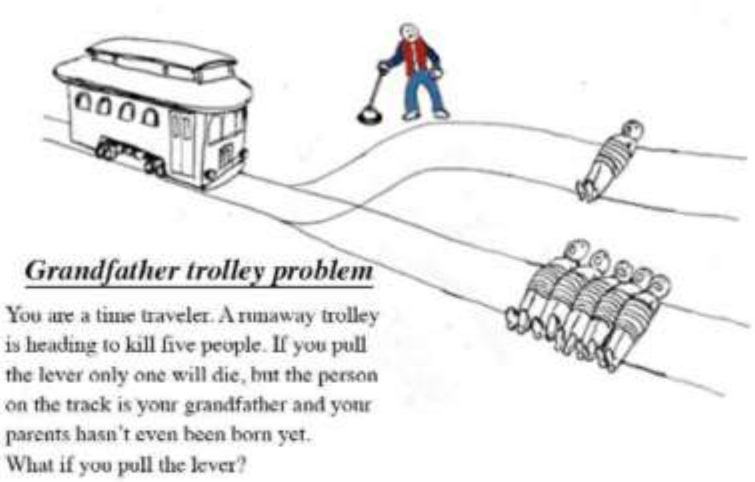
- Variation 3: The Fat Man Trolly Problem



https://themindcollection.com/trolley-problem-meme-variations/

What would you do?

- Variation 4: Grandfather Trolly Problem



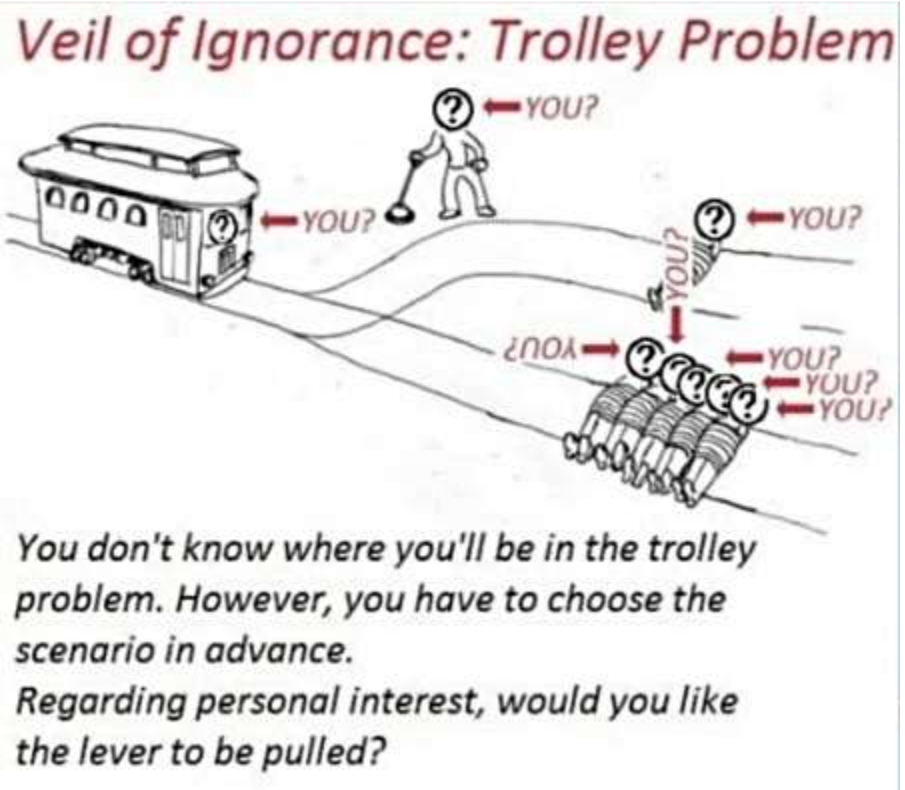
Grandfather trolley problem

You are a time traveler. A runaway trolley is heading to kill five people. If you pull the lever only one will die, but the person on the track is your grandfather and your parents hasn't even been born yet.
What if you pull the lever?

https://themindcollection.com/trolley-problem-meme-variations/

What would you do?

- Variation 5



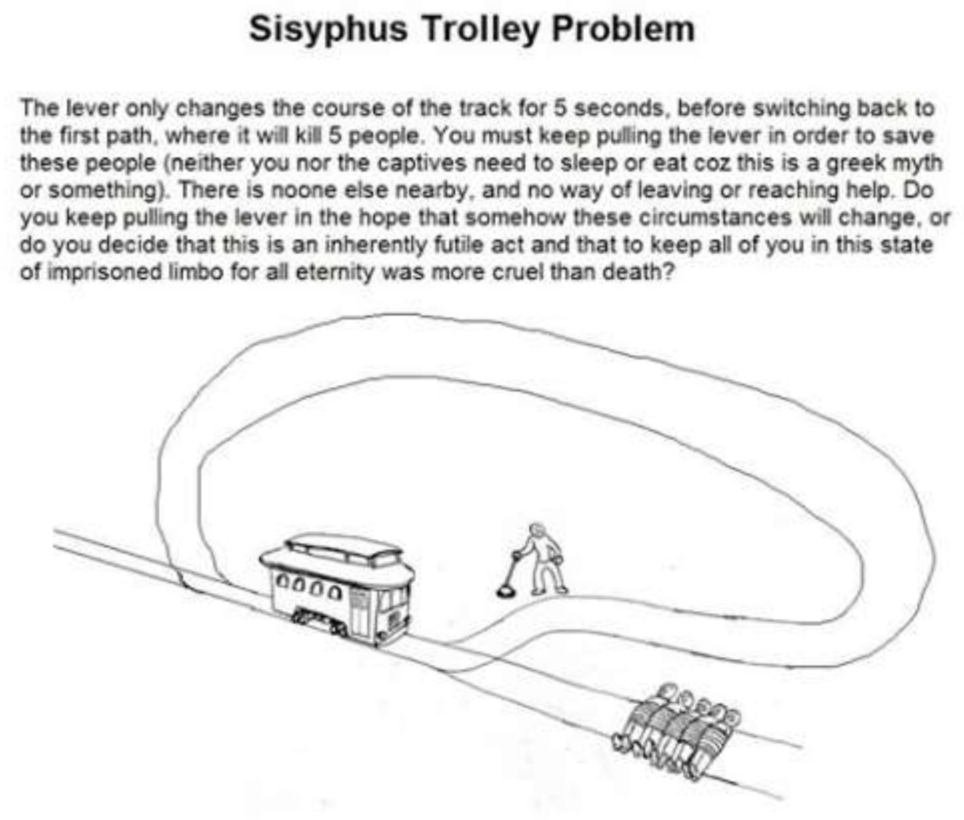
Veil of Ignorance: Trolley Problem

You don't know where you'll be in the trolley problem. However, you have to choose the scenario in advance.
Regarding personal interest, would you like the lever to be pulled?

https://themindcollection.com/trolley-problem-meme-variations/

What would you do?

- Variation 6

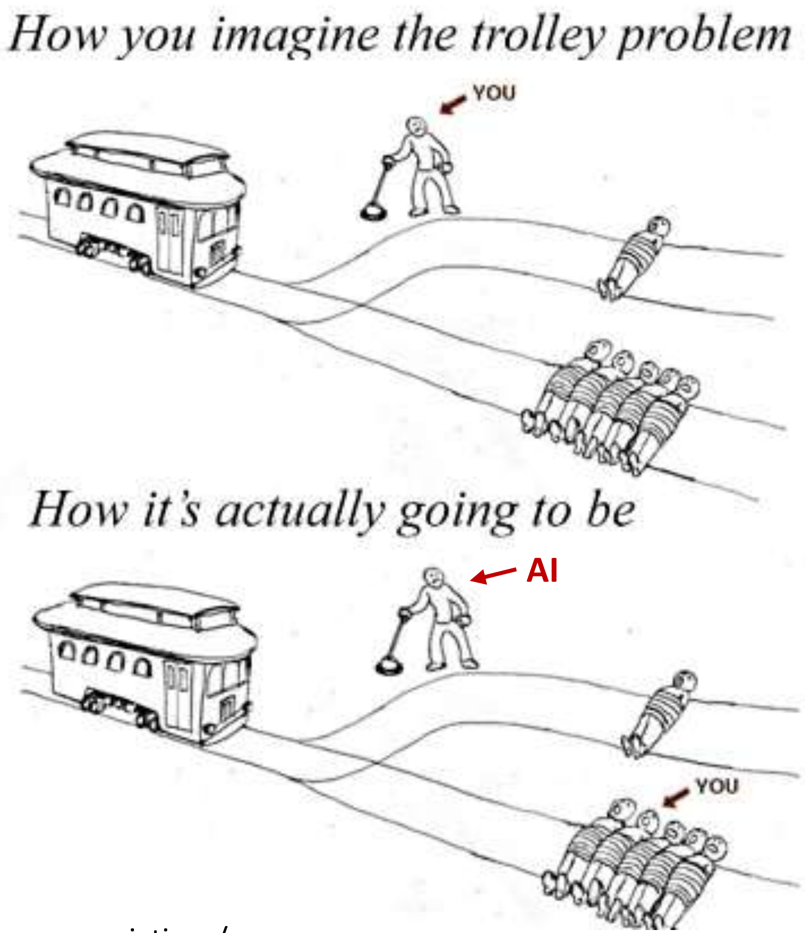


Sisyphus Trolley Problem

The lever only changes the course of the track for 5 seconds, before switching back to the first path, where it will kill 5 people. You must keep pulling the lever in order to save these people (neither you nor the captives need to sleep or eat coz this is a greek myth or something). There is noone else nearby, and no way of leaving or reaching help. Do you keep pulling the lever in the hope that somehow these circumstances will change, or do you decide that this is an inherently futile act and that to keep all of you in this state of imprisoned limbo for all eternity was more cruel than death?

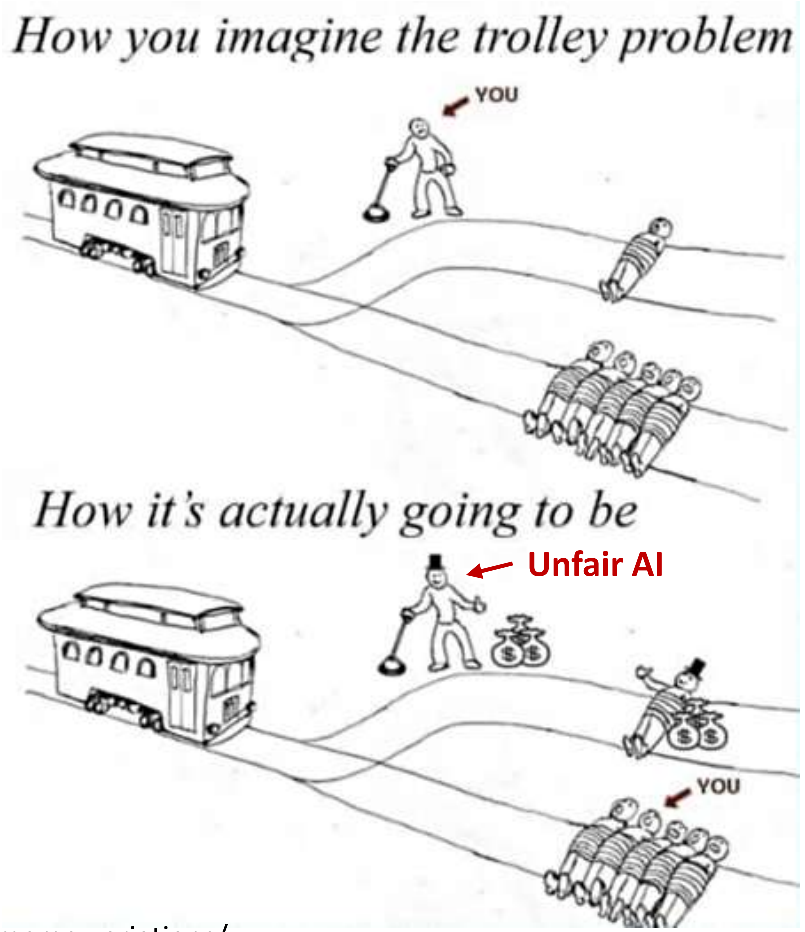
https://themindcollection.com/trolley-problem-meme-variations/

Actual Trolly Problem in the Age of AI


How you imagine the trolley problem
YOU
How it's actually going to be
AI
YOU

https://themindcollection.com/trolley-problem-meme-variations/

Trolly Problem Shows Challenges of AI Ethics


How you imagine the trolley problem
YOU
How it's actually going to be
Unfair AI
YOU

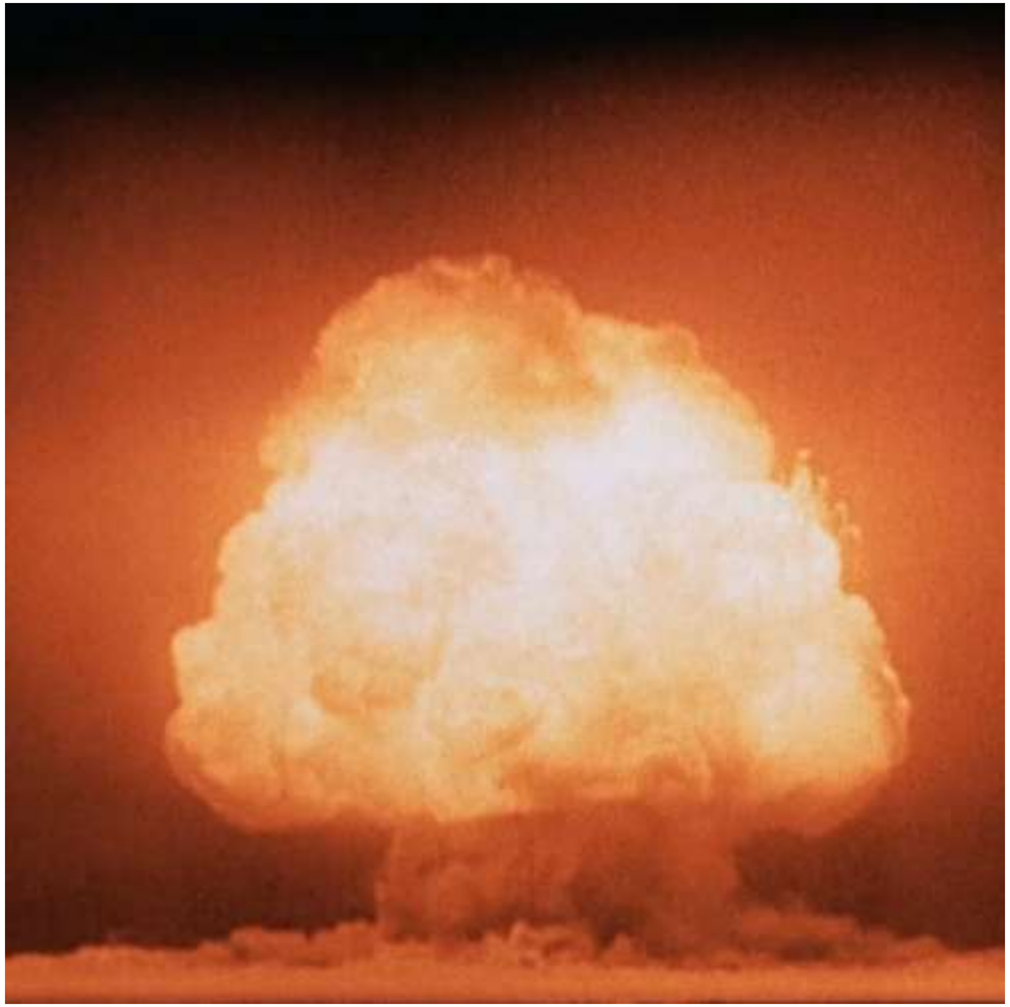
https://themindcollection.com/trolley-problem-meme-variations/

What should the AI (e.g., self-driving car) do?

- **Scenario:** Cargo fell off the truck ahead, and the brakes are insufficient to avoid a collision. The **self-driving car** faces three options: first, turn left and collide with an SUV; second, turn right and collide with a cyclist; third, go straight and crash into an obstacle ahead.
- Strategies:
  - 1. Prioritize self-safety.
  - 2. Minimize overall damage.
  - 3. Additional information.
  - 4. Random strategy.

July 16, 1945

- First nuclear detonation
- Alamogordo, NM
  - "Trinity" test
- Manhattan Project
- Fears over Nazi development of an atomic bomb
- Employed 130,000 people
- $2 billion dollars invested in project development



August 6, 1945
Hiroshima

August 9, 1945
Nagasaki

July 9, 1955

- “We appeal as human beings to human beings: Remember your humanity, and forget the rest. If you can do so, the way lies open to a new Paradise; if you cannot, there lies before you the risk of universal death.”

  ---Russell-Einstein Manifesto

- Signed by Max Born (1954), Percy Bridgman (1946), Leopold Infeld, Frederic Joliot-Curie (1935), Herman Muller (1946), Linus Pauling (1954), Cecil Powell (1950), Joseph Rotblat, and Hideki Yukawa (1949)
- Called Congress & scientists to assemble and discuss the threat posed by thermonuclear weapons

1957

- Pugwash, Nova Scotia
- Attendance of 22 international scientists
  - 7 from USA
  - 3 from Japan, USSR
  - 2 from UK, Canada
  - 1 from Australia, China, France, Poland
- Pugwash Conference on Science and World Affairs
- Founders: Rotblat & Russell
- 1995 – awarded the Nobel Peace Prize (with Rotblat)
  - "for their efforts to diminish the part played by nuclear arms in international politics and, in the longer run, to eliminate such arms".

Bertrand Russell

Ethical Reasoning

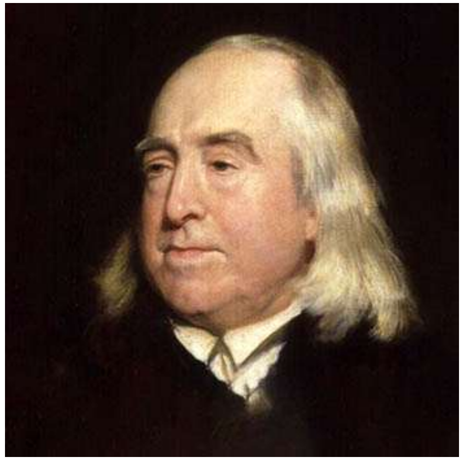
- Consequentialist (结果论者)
  - Choose the action with the best overall consequences
  - Morally right if the act, or motive, is chosen only with consideration for the consequence
  - “The ends justify the means”
  - Discounts the treatment of people and actions
- Obligationist (义务论者)
  - Certain actions are inherently right or wrong
  - Do not balance the consequences to determine obligations

Some Ethical Codes

Code	Model	Type
Utilitarianism (功利主义)	Choose action that maximizes the good for the greatest number of people	Consequentialist
Universal Application Code	Action is “right” if everyone accepts the moral rule presupposed by the action	Obligationist
Golden Rule	Reciprocity - “Do unto others as you’d have done to you”	Obligationist
Human Rights	Respect the rights of others (i.e., right to life, safety, privacy, etc.)	Obligationist
Natural Law	Action is “right” if legitimate authority says it is (legislative body, professional society) “right”	Obligationist

Utilitarianism

- Dates to Epicurus (~300 BC)
- Formalized by Bentham
  - “The greatest happiness principle” (a.k.a. “the principle of utility”)
    - Term attributed to David Hume
  - Evaluation of actions based on their consequences
  - Actions that do not maximize the greatest happiness are considered as morally wrong
- Popularized by Mill
  - Actions rooted in self-interest
  - Personal feelings often conflict with self-interest, but they not contrary to pleasure



Jeremy Bentham
(1748-1832)

John Stuart Mill
(1806-1873)

Utilitarianism - Problems

- If everyone acts in pure self-interest, there is a potential for suboptimal results
- Tragedy of the Commons（公地悲剧）
  - Limited resources
  - Each action has a positive and negative component
    - Positive – personal benefit by consuming resource
    - Negative – less resource available
  - At any particular point in time, personal gain outweighs the personal share of distributed cost
  - Overuse of resources

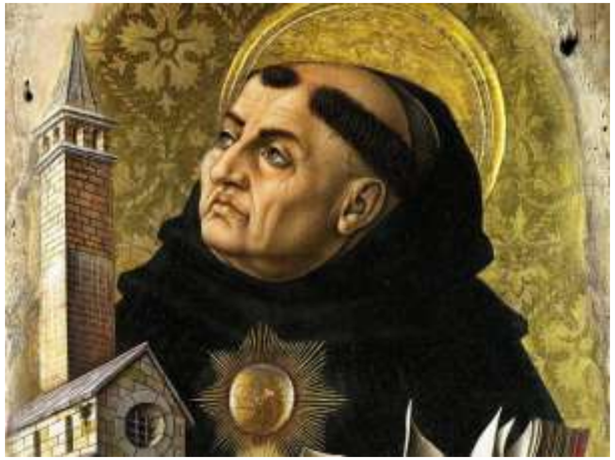
G. Hardin. The tragedy of the commons. Science. 1968; 162: 1243-1248.

Utilitarianism Relaxed

- Some people must lose so that others may benefit
- Challenge is in defining “utility”
    - Everyone may have a different notion of what is “best”

M. McGuire. The calculus of moral obligation. Ethics. 1985; 95(2): 199-223.

Principle of Double Effect

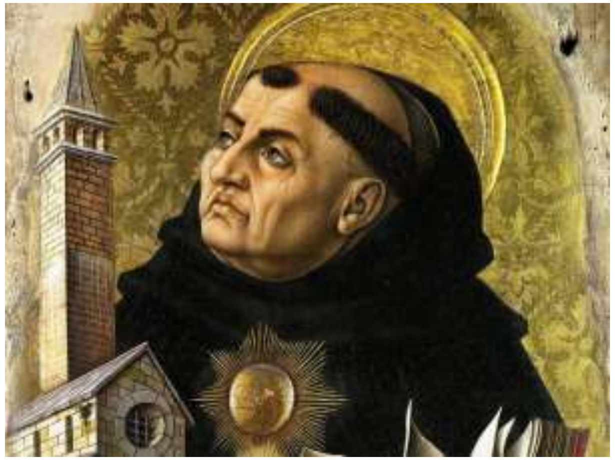


St. Thomas Aquinas
(1225-1275)

- A.k.a. Doctrine of Double Effect
- Attributed to St. Thomas Aquinas
  - Italian Catholic philosopher and theologian
- System of ethics
  - Rooted in Aristotle reasoning
  - Will wills the end, it also wills the appropriate means – chosen freely
- Man’s will is an inclination toward universal good
- In a chain of acts, man strives toward the highest end
- Free acts, such that man has knowledge of their end
- If an act is “good” or “evil” depends on the end

Principle of Double Effect

- Certain actions have multiple effects
- Special case – two effects
  - Good effect: Result is positive benefit for the recipient
  - Bad effect: Result is a negative cost for the recipient
  - Do not necessarily observe both the good and bad effects



St. Thomas Aquinas
(1225-1275)

Principle of Double Effect

- Four (sometimes five) conditions

1. The act’s nature is morally neutral (or positive)
2. Intent of the actor is for the good, not the bad, effect
3. Good effect outweighs the bad effect in a situation sufficiently grave to merit the risk of generating the bad effect
4. Good effect does not proceed through the bad effect
5. From a causal perspective, the bad effect is no closer to the act than the good effect

Summary of Principle of Double Effect

- The entity initiating the action can not inflict harm to achieve good.

- Note, this does not imply there are no harms that manifest as a bi-product.

Moral Neutrality

- An entity or object that has no intent regarding its action
- A tool, such as screwdriver, can be used in either a morally defensible or morally problematic way
- If a pillowcase is used by a perpetrator for suffocation, the pillowcase did not intend the harm.
- That said, are all **AI** technologies morally neutral? （技术无罪论）

Recap of AI Models

Markov Models

- Markov models are probabilistic models that capture dependencies between events.

Neural Network Models

- Neural network models are computational models inspired by the human brain.

Generative AI

- Generative AI combines the concepts of Markov and neural network models.
- It leverages the probabilistic nature of Markov models and the learning ability of neural network models to generate new data or content.

Brief History of AI Development

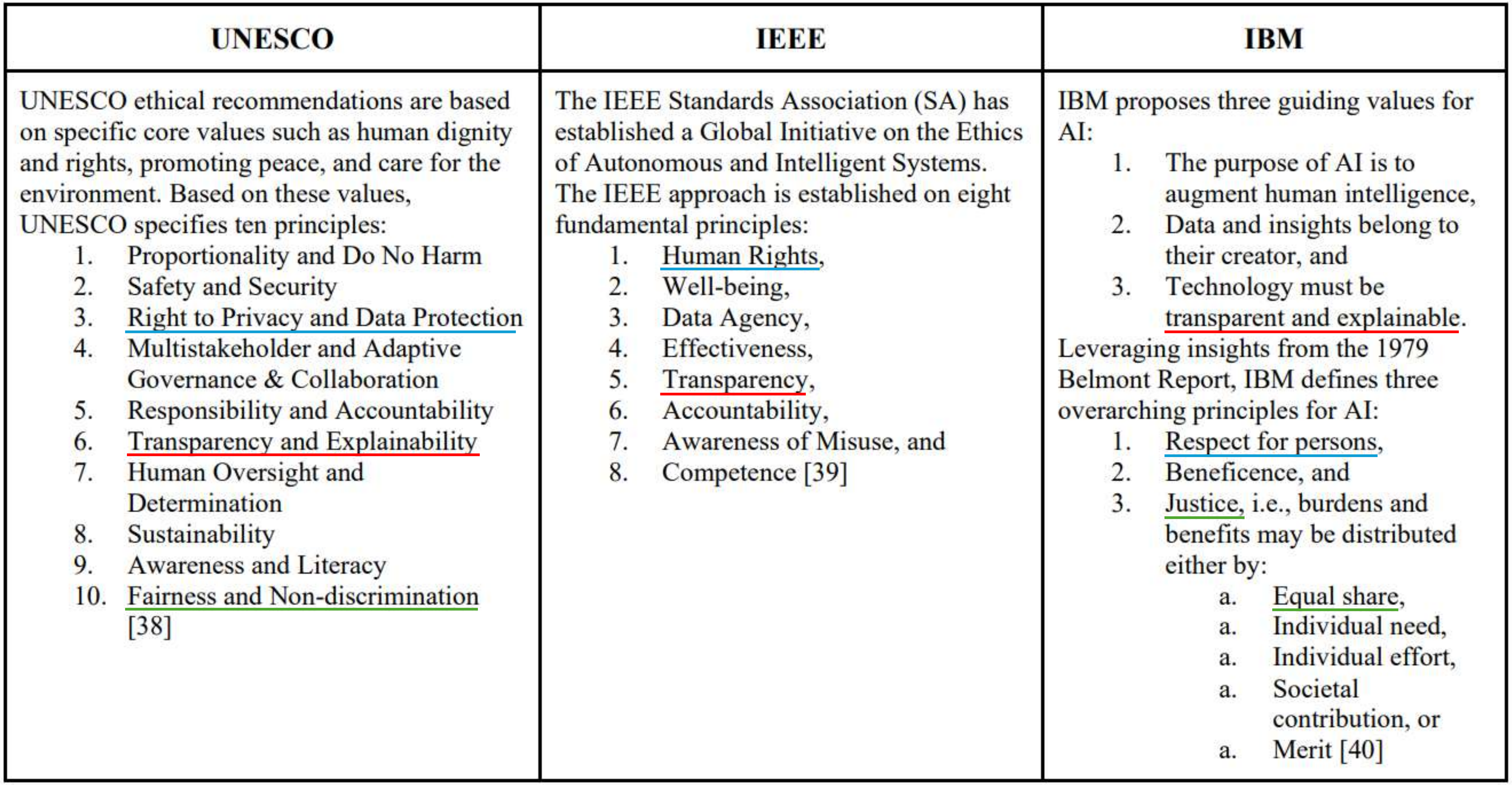
17th Century	Pascal and Leibniz	Ideas of Intelligent Machines
1920s	Charles Babage	1st "Computing Machine"
1950	Alan Turing	Proposed "Turing Test"
1955-1956	John McCarthy	Organized the Dartmouth Conference, proposed the concept of AI
1997	IBM	Deep Blue (Won Chess World Champion)
1990s	Vapnik and Chervonenkis	SVM
1998	Yann LeCun	LeNet-5 -> CNN
2006	Hinton	Neural networks
2012	Hinton et al.	AlexNet
2014	Ian Goodfellow	Generative adversarial network (GAN)
2015	Kaiming He et al.	ResNet
2017	Google DeepMind	AlphaGo (Won Go Chess World Champion)
2018	Google AI	Bidirectional Encoder Representations from Transformers (BERT)
2021	Google DeepMind	AlphaFold
2022	OpenAI	ChatGPT-3.5 (ChatGPT based on GPT-3.5)
2023	OpenAI	ChatGPT-4 (ChatGPT based on GPT-4)

Group Discussion

- In a group of 2-3, discuss the following questions:
  - What should an ethical AI system look like?
  - What properties and principles should it have?
  - Can you give several examples of ethical AI systems?
  - Can you give several examples of unethical AI systems?

Responsible AI ≈ Ethical AI

- **Fairness**: AI systems should be unbiased and not discriminate against any group of people
- **Privacy**: AI systems should protect people's personal information
- **Transparency**: AI systems should be clear and understandable
- **Accountability**: Organizations should be held accountable for how they use AI
- **Non-maleficence**: AI systems should not harm individuals, society, or the environment
- **Inclusiveness**: AI systems should engage with diverse perspectives



UNESCO	IEEE	IBM
UNESCO ethical recommendations are based on specific core values such as human dignity and rights, promoting peace, and care for the environment. Based on these values, UNESCO specifies ten principles: 1. Proportionality and Do No Harm 2. Safety and Security 3. Right to Privacy and Data Protection 4. Multistakeholder and Adaptive Governance & Collaboration 5. Responsibility and Accountability 6. Transparency and Explainability 7. Human Oversight and Determination 8. Sustainability 9. Awareness and Literacy 10. Fairness and Non-discrimination [38]	The IEEE Standards Association (SA) has established a Global Initiative on the Ethics of Autonomous and Intelligent Systems. The IEEE approach is established on eight fundamental principles: 1. Human Rights, 2. Well-being, 3. Data Agency, 4. Effectiveness, 5. Transparency, 6. Accountability, 7. Awareness of Misuse, and 8. Competence [39]	IBM proposes three guiding values for AI: 1. The purpose of AI is to augment human intelligence, 2. Data and insights belong to their creator, and 3. Technology must be transparent and explainable. Leveraging insights from the 1979 Belmont Report, IBM defines three overarching principles for AI: 1. Respect for persons, 2. Beneficence, and 3. Justice, i.e., burdens and benefits may be distributed either by: a. Equal share, a. Individual need, a. Individual effort, a. Societal contribution, or a. Merit [40]

Table 1. Ethical Principles Statements from selected organizations

Kirova VD, Ku CS, Laracy JR, Marlowe TJ. The ethics of artificial intelligence in the era of generative AI. Journal of Systemics, Cybernetics and Informatics. 2023 Dec;21(4):42-50.

Case studies

- What types of data were misused? What principles of ethical AI were violated? What AI model were used? How to mitigate the risks?

Case 1	In 2018, Amazon's facial recognition system misidentified 28 lawmakers as criminals.
Case 2	In 2017, Vietnamese security company, Bkav, used 3D-printed mask to bypass iPhone's Face ID.
Case 3	In 2019, Criminals used AI technology to mimic CEO's voice.
Case 4	UK passport photo AI verification system shows bias against Black women.
Case 5	Amazon's voice assistant Alexa recommends a 10-year-old girl to use a coin to touch a socket.
Case 6	YouTube's recommendation algorithm suggests inappropriate videos to children.
Case 7	Ride-hailing platforms recommend different car models based on the user's phone brand and price.
Case 8	In 2020, Tesla's self-driving car failed to recognize a white truck, leading to an accident.
Case 9	In 2018, Uber's self-driving vehicle stroked a pedestrian at night.

Group Discussion

- In a group of 2-3, discuss the following questions:
  - There are now various methods of identity recognition, including facial recognition, voice recognition, and fingerprint recognition, with related AI technologies becoming increasingly mature. However, data such as facial photos and voice recordings are easily accessible, posing significant security risks. Can you think of some other identity recognition methods with higher security?
  - Long-term use of recommendation systems can easily lead to the “filter bubble” effect. How can users avoid falling into this “filter bubble”? Do you prefer the system to recommend the most interesting information to you, or would you prefer it to recommend information from different fields?
  - What ethical issues exist in robot care for the elderly? How to solve?
  - What ethical issues exist in Brain Computer Interface? How to solve?

Brainstorming

- “In 15 years, AI and automation will have the technological capability to replace 40% of jobs.” – Kaifu Li said in June, 2016

Jobs will be replaced by AI in 5 years (2030)	Jobs will NOT be replaced by AI in 5 years (2030)
	• Medical researcher, artificial intelligence scientist, screenwriter, public relations expert, entrepreneur. • CEO, negotiation expert, mergers and acquisitions expert. • Oral surgeon, aircraft mechanic, chiropractor. • Geological survey cleaner. • Social worker, special education teacher, marriage counselor.

Readings Due on October 11

- Murdoch B. **Privacy and artificial intelligence: challenges for protecting health information in a new era**. *BMC medical ethics*. 2021 Sep 15;22(1):122.
  - ❑ https://link.springer.com/article/10.1186/s12910-021-00687-3
- Optional
  - ❑《工程伦理》Ch.10.
  - ❑《信息科学技术伦理与道德》Chs.5&7.
  - ❑Price WN, Cohen IG. **Privacy in the age of medical big data**. *Nature medicine*. 2019 Jan;25(1):37-43.
  - ❑ S. Warren and L. Brandeis. **The right to privacy**. *Harvard Law Review*. 1890; V. IV, No. 5.
    - http://faculty.uml.edu/sgallagher/Brandeisprivacy.htm