

Medical Data Privacy and Ethics in the Age of Artificial Intelligence

Lecture 2: AI Ethics

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

Assistant Professor of Biomedical Engineering

ShanghaiTech University

September 23, 2026

以下文章来源于智能涌现，作者温丽虹

智能涌现

直击AI新时代下涌现的产业革命。36氪旗下账号。



文 | 温丽虹

编辑 | 张雨忻

来源 | 智能涌现 (ID: AIEmergence)

封面来源 | 受访者提供

9月份，导演陆川⁹让AI完全接管了自己的一台电脑。他移除了电脑上所有敏感、保密的信息，训练AI使用那台电脑上的剪辑软件、三维制作软件与后期制作软件。他计划着，要在那台电脑上做一个实验，让AI从一部一分钟片子开始，尝试自主创作。

在陆川的定义里，制作和创作是两件事。

制作，AI可以做得很好，甚至超出大多数人的想象。“AI出手，基本能做到80分。因为它汲取了——或者说盗取了——人类的火种。学习过人类历史上几乎所有优秀的超然的作品。”

但创作不同。创作由欲望催生，“AI到现在为止没有任何科学依据证明它有欲望。它没有恐惧，没有对死亡的恐惧，没有对生的欲求，没有对自由的向往，也没有爱。”说着这话，陆川自顾自陷入了思考。“创作的本质是不完美，是性格的缺陷，你精神上一定有缺陷，灵魂中一定有黑洞，情感上一定是一种痛苦。人一定有绝望，才能发出呐喊。”

Meta used AI to target workers with medical conditions for layoffs, lawsuit claims

By Daniel Wiessner

July 14, 2026 9:52 PM GMT+8 · Updated July 15, 2026



A woman walks by the Meta Lab in Los Angeles, California, U.S., May 20, 2026. REUTERS/Daniel Cole [Purchase Licensing Rights](#)

Summary	Companies
- The 26 plaintiffs seek a court order blocking layoffs set to start on July 22 - Meta said the claims lack merit and people made workforce decisions, not AI - Lawsuit says Meta used productivity scores and AI token usage in layoff selections	

July 14 (Reuters) - Twenty-six employees of Meta Platforms (META.O) have filed a novel lawsuit accusing the tech giant of using AI-powered software that disproportionately targeted people with disabilities or who took medical leave in selecting workers for mass layoffs.

The lawsuit, filed in Oakland, California, federal court late Monday, says that the company relied on factors such as productivity and AI token usage when it slashed thousands of jobs earlier this year, disadvantaging people who missed work because of medical conditions or to care for family members.



The UK Biobank breach shows the need for data management training

Modern research culture's encouragement of widespread sharing makes disclosure risks inevitable, say Elizabeth Green and Felix Ritchie

Published on May 25, 2026

Last updated May 25, 2026

Elizabeth Green Felix Ritchie

Universities increasingly invoke the language of “trusted” and “secure” research. Yet many institutions using these terms do not fully understand what those concepts operationally require. That is why the recent controversies surrounding UK Biobank cannot simply be put down to a few “bad apples” or isolated failures of individual judgement.

Last month, researchers were temporarily blocked from accessing the database of anonymised medical information on more than 500,000 UK volunteers after some of that information was offered for sale online in China.

But while the incident constituted a shocking breach of research ethics, it is equally true that we should have seen this coming.

Most academics receive extensive training in ethics, informed consent and research integrity. But far fewer receive practical training in how to safely manage and share data; on the contrary, modern research culture encourages openness, reproducibility and the widespread sharing of data and code, in ways that make disclosure risks inevitable.

The UK has long been a global leader in managing confidential research data. National statistical agencies, government departments and specialist data services – the UK Data Archive and the Office for National Statistics among them – have spent decades developing systems to enable research access while protecting subjects’ identities.

Central to this is the Five Safes framework. This recognises that safe research depends on multiple interacting safeguards related to the project in question, the people involved, the security of the data analysis environment, the extent to which the data are anonymised and the project’s public outputs.



Source: Black_Kira/iStock

Goals of this lecture

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

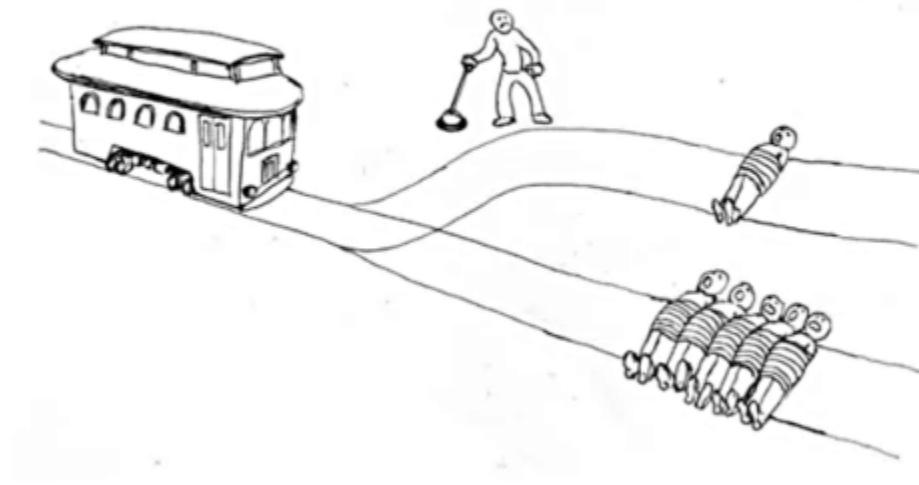
Introduction to Ethics

- **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?

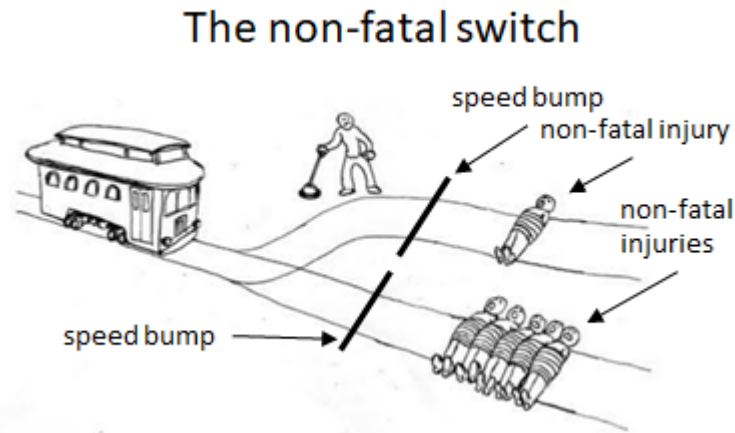
• Trolley Problem

- 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?

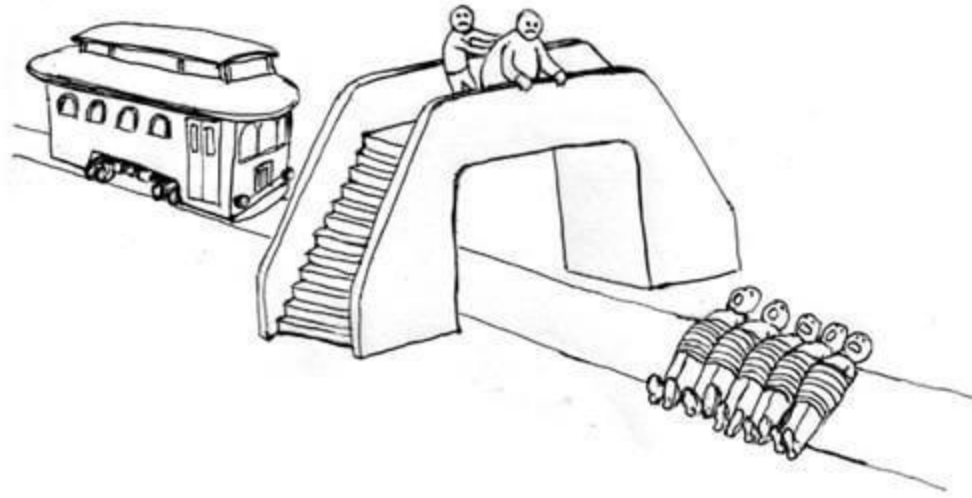
- Variation 1: The Non-fatal Switch



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

What would you do?

- Variation 2: The Fat Man Problem

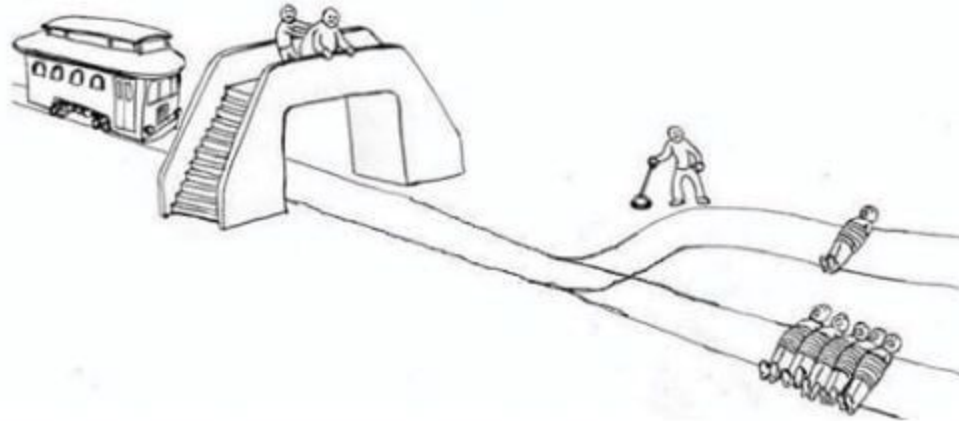


<https://themindcollection.com/trolley-problem-meme-variations/>

What would you do?

- Variation 3: The Fat Man Trolley Problem

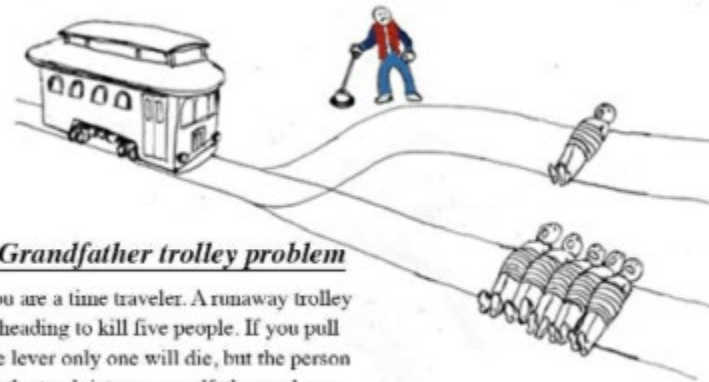
Do you push the fat man to prevent a trolley problem?



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

What would you do?

- Variation 4: Grandfather Trolley Problem



Grandfather trolley problem

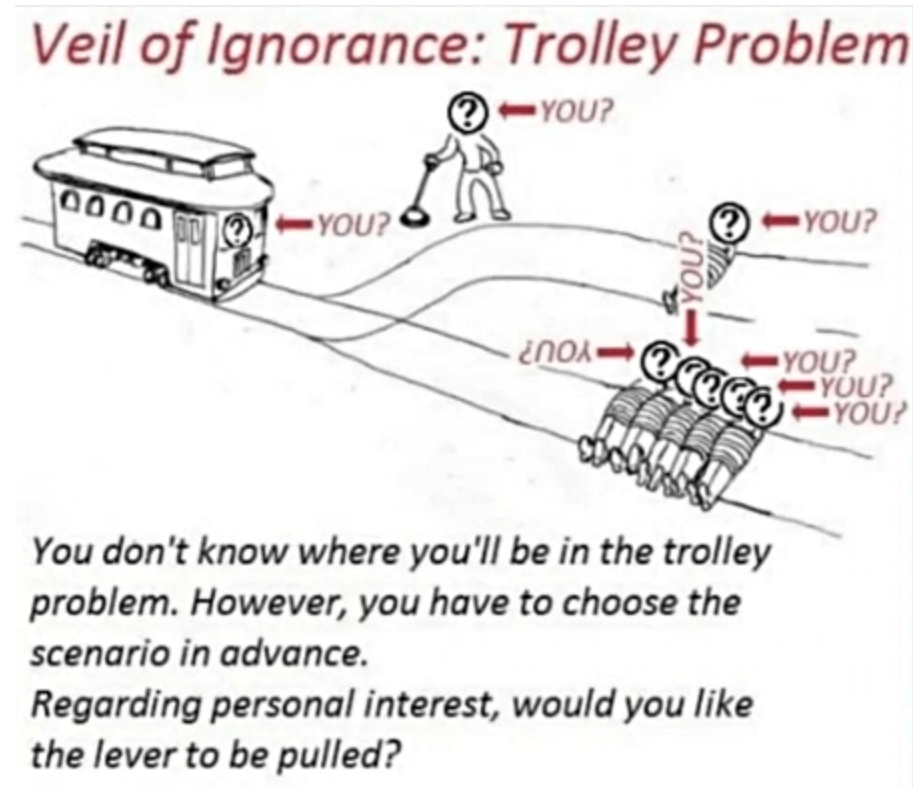
You are a time traveler. A runaway trolley is heading to kill five people. If you pull the lever only one will die, but the person on the track is your grandfather and your parents hasn't even been born yet. What if you pull the lever?



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

What would you do?

- Variation 5



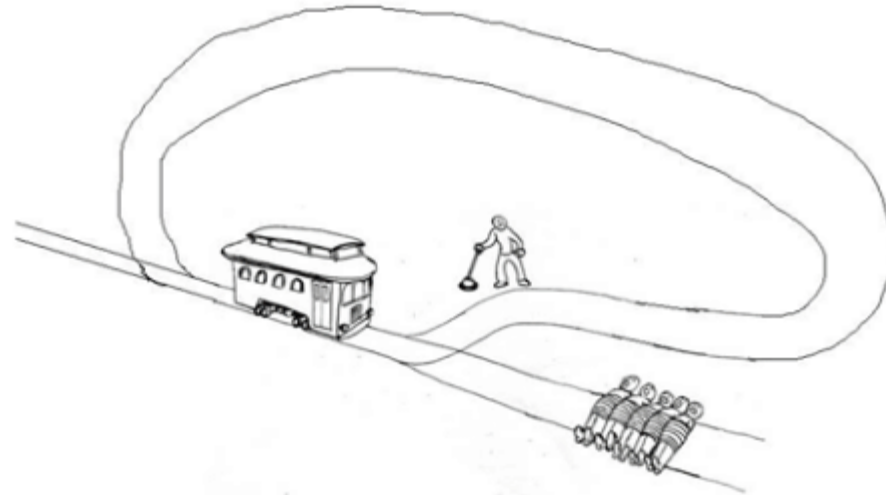
<https://themindcollection.com/trolley-problem-meme-variations/>

What would you do?

- Variation 6

Sisyphus Trolley Problem

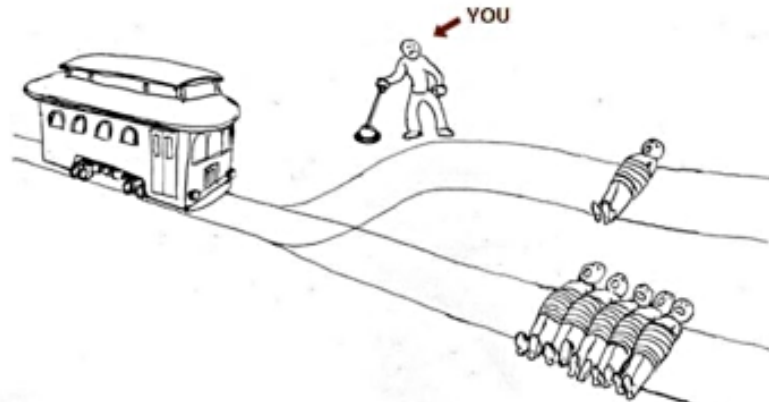
The lever only changes the course of the track for 5 seconds, before switching back to the first path, where it will kill 5 people. You must keep pulling the lever in order to save these people (neither you nor the captives need to sleep or eat coz this is a greek myth or something). There is noone else 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?



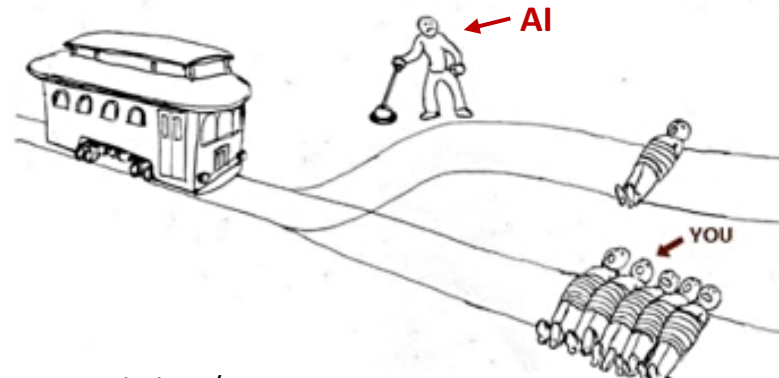
<https://themindcollection.com/trolley-problem-meme-variations/>

Actual Trolley Problem in the Age of AI

How you imagine the trolley problem

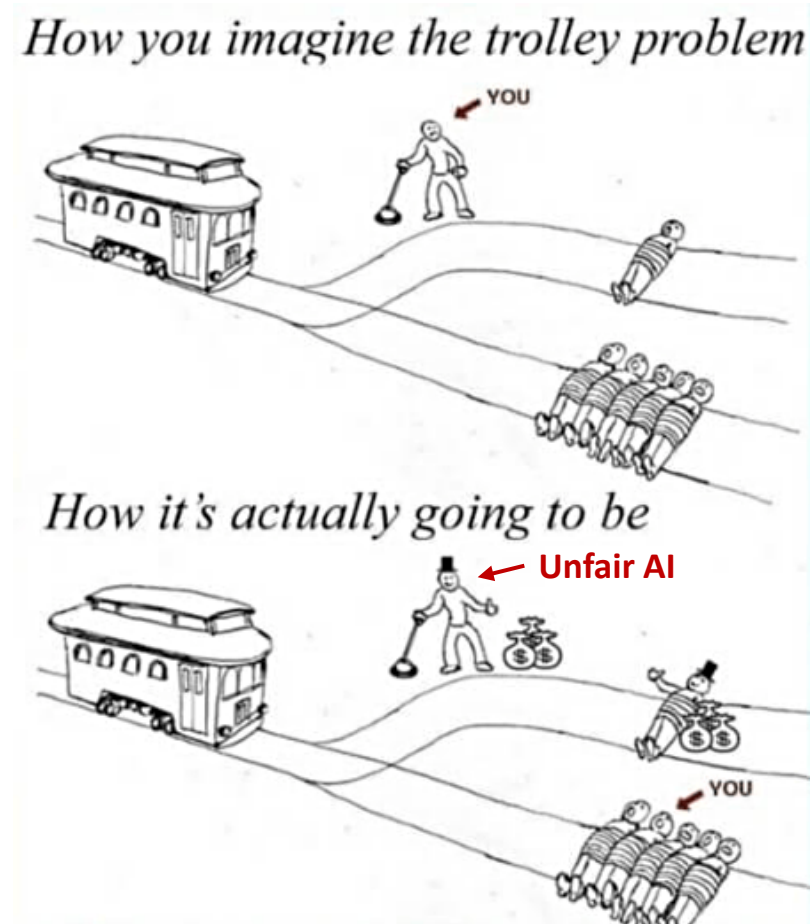


How it's actually going to be



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

Trolley Problem Shows Challenges of AI Ethics



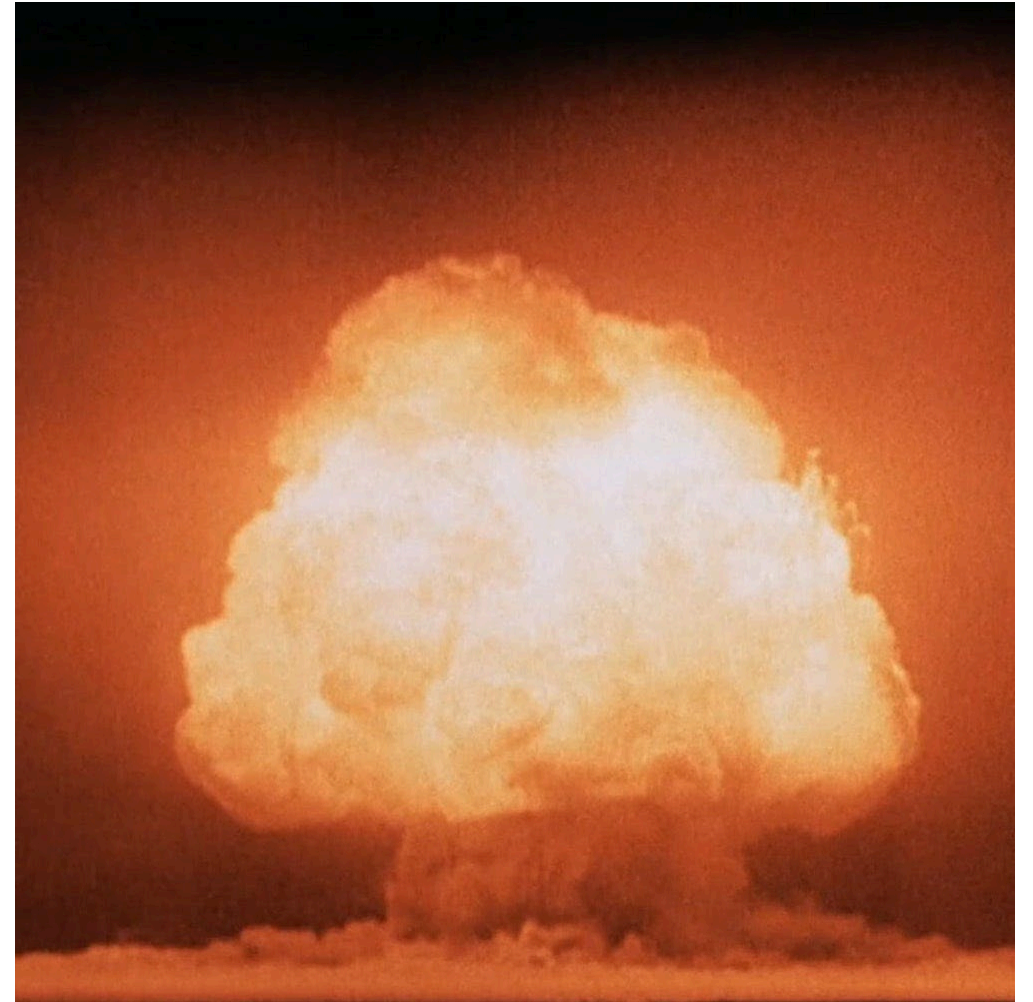
<https://themindcollection.com/trolley-problem-meme-variations/>

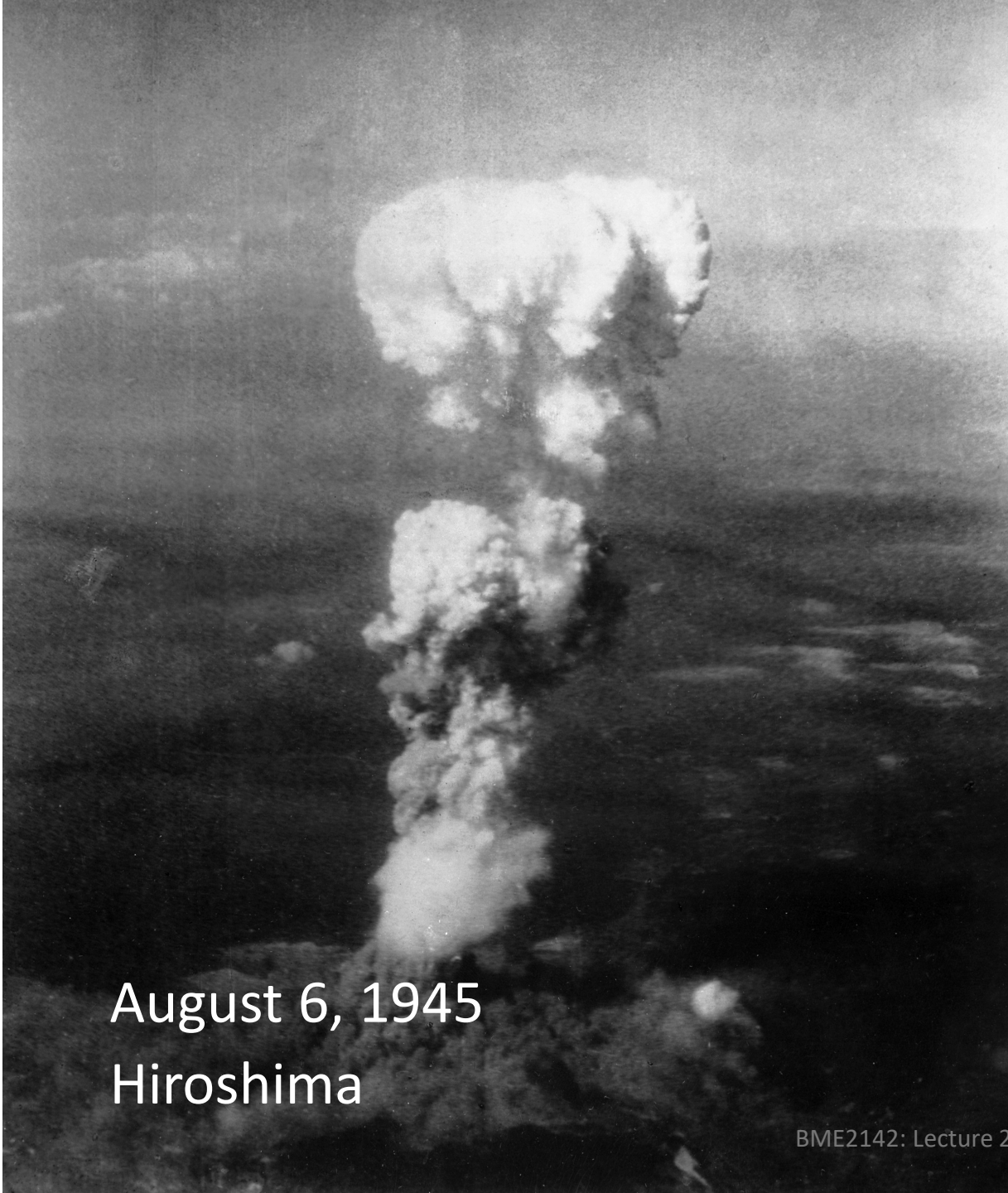
What should the AI (e.g., self-driving car) do?

- **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

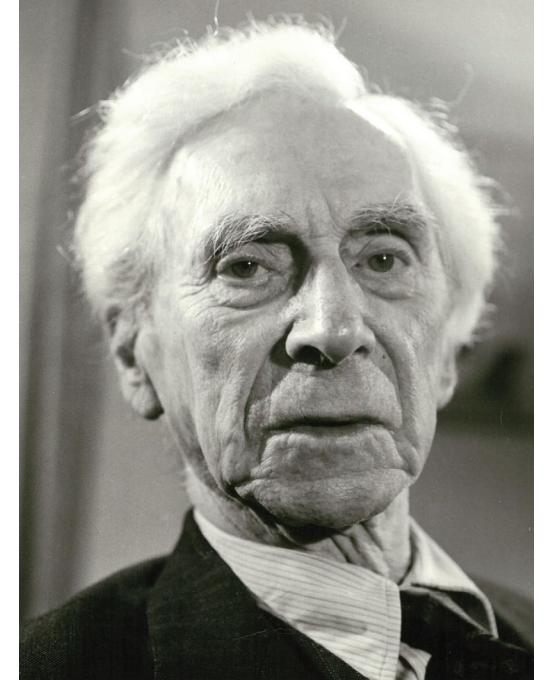
- “We appeal as human beings to human beings: Remember your humanity, and forget the rest. If you can do so, the way lies open to a new Paradise; if you cannot, there lies before you the risk of universal death.”

---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

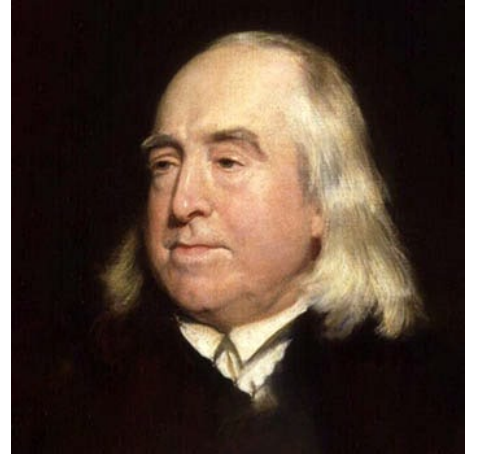
- Consequentialist (结果论者)
 - Choose the action with the best overall consequences
 - Morally right if the act, or motive, is chosen only with consideration for the consequence
 - “The ends justify the means”
 - Discounts the treatment of people and actions
- Obligationist (义务论者)
 - Certain actions are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

Some Ethical Codes

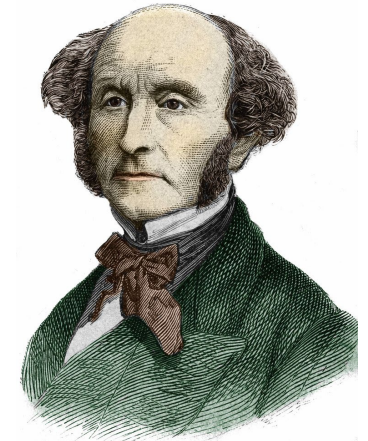
Code	Model	Type
Utilitarianism (功利主义)	Choose action that maximizes the good for the greatest number of people	Consequentialist
Universal Application Code	Action is “right” if everyone accepts the moral rule presupposed by the action	Obligationist
Golden Rule	Reciprocity - “Do unto others as you’d have 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”

Principle of Double Effect

- 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



St. Thomas Aquinas
(1225-1275)

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)

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?

Principles of AI Ethics

- **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. <u>Right to Privacy and Data Protection</u> 4. Multistakeholder and Adaptive Governance & Collaboration 5. Responsibility and Accountability 6. <u>Transparency and Explainability</u> 7. Human Oversight and Determination 8. Sustainability 9. Awareness and Literacy 10. <u>Fairness and Non-discrimination</u> [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. <u>Human Rights</u>, 2. Well-being, 3. Data Agency, 4. Effectiveness, 5. <u>Transparency</u>, 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 <u>transparent and explainable</u>. Leveraging insights from the 1979 Belmont Report, IBM defines three overarching principles for AI: 1. <u>Respect for persons</u>, 2. Beneficence, and 3. <u>Justice</u>, i.e., burdens and benefits may be distributed either by: a. <u>Equal share</u>, 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 about AI Ethics

- 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?

Readings Due on Sept 30

■ Optional

- ❑ 《工程伦理》 Chs.10&12.
- ❑ 《信息科学技术伦理与道德》 Chs.5.
- ❑ 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>