Ethics in AI

Dr Yuanwei Liu
Associate Professor, IEEE Fellow
Web of Science Highly Cited Researcher,
IEEE Communication Society Distinguished Lecturer,
IEEE Vehicular Technology Distinguished Lecturer,
e-mail: ***yuanwei.liu@qmul.ac.uk***
website: ***http://www.eecs.qmul.ac.uk/~yuanwei/***

Queen Mary
University of London

Outline

- Four Mainstream Categories of AI Ethical Issues
- AI Ethical Issues Summary
- Ethical Guidelines and Principles for AI
- IEEE Guidelines of Ethics
- Case Study: AI Ethics in Ride Sharing

AI poses many significant ethical risks or issues

- In 2016, the driver of an **electric Tesla car** was killed in a road accident after its Autopilot mode failed to recognize an oncoming lorry

- Microsoft's AI chatting bot, **Tay.ai**, was taken down because it became racist and sexist less than only a day after she joined Twitter

- *Other examples concerned with failure, fairness, bias, privacy...*

- More seriously, AI technology has begun to be used by criminals to harm others or society.
- For example, criminals used AI-based software to impersonate a chief executive's voice and demand a **fraudulent transfer of $243, 000**.

Recognize and understand the potential AI ethical issues or risks

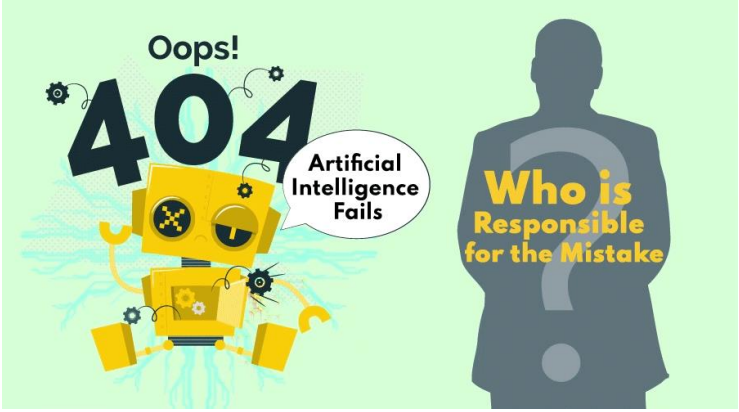
Four different categorizations of AI ethical issues

1. Categorization Based on Features of AI, Human Factors and Social Impact

➢ Ethical issues caused by **features of AI**

- **Transparency**: It is hard to explain and understand the inference procedure of ML, which is commonly known as the "black box".
- **Data Security and Privacy**: The misuse and malicious use of data, such as (personal) information leakage or tampering, are serious ethical issues that are closely related to every individual, institution, organization, and even the country.
- **Autonomy, Intentionality, and Responsibility**: How much autonomy, intentionality, and responsibility should an AI system be allowed is a challenging question and issue.

Recognize and understand the potential AI ethical issues or risks



➢ Ethical issues caused by **human factors**

- **Accountability**: When an AI system or agent fails in a specified task and results in bad consequences, who should be responsible.
- **Ethical Standards**: As the ultimate goal of AI ethics is to create ethical AI that can follow ethical principles and behave ethically, it is crucial to form comprehensive and unbiased ethical standards for training or regulating AI to be ethical.
- **Human Rights Laws**: The designer, software engineers, and other participants in AI system design and application should be taught human rights laws.

Recognize and understand the potential AI ethical issues or risks

➢ **Social impact** of ethical AI issues

- **Automation and Job Replacement**: As more and more factory workers are being replaced by automated systems and robots, AI will disrupt and transform the labor market.

- **Accessibility**: The accessibility or availability of emerging technologies, such as AI, will have a direct impact on human well-being. However, it will be unethical and unfair if only a portion of the population benefit from AI.

- **Democracy and Civil Rights**: Unethical AI will distort the truth and eventually lead to the loss of trust and public support for AI technology.

Recognize and understand the potential AI ethical issues or risks

Four different categorizations of AI ethical issues

2. Categorization Based on Vulnerabilities of AI and Human

➢ Ethical issues from the **vulnerabilities of AI**

- **ML is data hungry**: Usually, ML requires a large amount of data to work well.
- **Garbage in/garbage out**: The performance of a ML algorithm heavily depends on the data from which it learns.
- **Faulty algorithms**: Even if a ML algorithm is input with enough and accurate data, if the algorithm itself is bad, it will also make bad predictions.
- **Deep learning is a black box:** Deep learning is a black box, which raises issues such as explainability, interpretability, and trust.

Recognize and understand the potential AI ethical issues or risks

➢ Ethical issues from the **vulnerabilities of human**

- **Abuse of AI**: AI technologies, such as facial recognition and image generation, can work better than humans. However, ethical issues exist because people may be tempted to use them for ill.

  --- For instance, a government could use facial recognition technology to monitor its citizens, and ML can be used to fabricate photos or videos so realistic that humans cannot tell that they are fake

- **Job replacement**: Since intelligent robots can perform certain tasks faster and better than humans, many people worry that robots and other AI technologies will replace a large part of current human labor in the near future.

- **Issues about robotic companions:** As AI robots become more and more sophisticated, they have begun to be regarded as companions of humans. This raises some ethical issues about the relationship between human and robotic companions

Recognize and understand the potential AI ethical issues or risks

Four different categorizations of AI ethical issues

3. Categorization Based on Algorithm, Data, Application, and Long-Term and Indirect Ethical Risks

➢ Ethical issues related to **algorithms**

- **Algorithm security**: 1) There is a risk of algorithm or model leakage; 2) The parameters of the AI algorithm model may be modified illegally by an attacker, which will cause the performance deterioration of the AI model and may lead to undesirable consequences.

- **Algorithm explainability**: Due to the black-box characteristic of many ML algorithms, the decision process of AI algorithms is hard to understand.

- **Algorithmic decision dilemma**: Even though we have designed an AI model well, we cannot foresee or predict the decisions of the algorithm and the consequence it will produce.

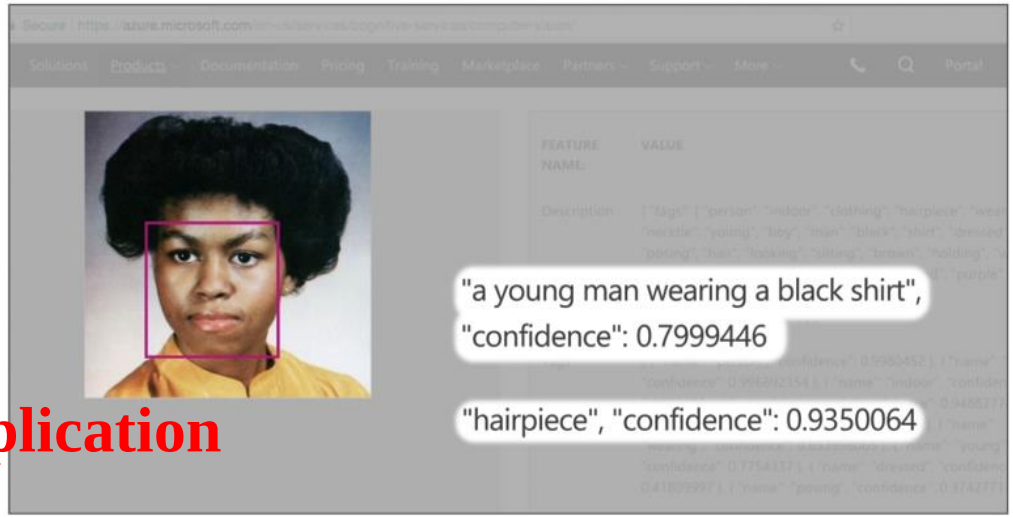
Recognize and understand the potential AI ethical issues or risks

➢ Ethical issues related to **data**

- **Privacy protection**: With the development of big data and AI, criminals have more ways to obtain personal privacy data with lower costs and greater benefits

- **Recognizing and processing personal and sensitive information**: If the personal or sensitive information is deidentified through randomization, data synthesis, and other technologies, it will no longer be regarded as personal or sensitive information and not protected by traditional laws.

Recognize and understand the potential AI ethical issues or risks



➢ Ethical issues related to **application**

- **Algorithm discrimination**: Algorithm discrimination or bias has been seen in many applications of AI.
  --- For instance, the racial bias in criminal justice systems and gender discrimination in hiring.

- **Algorithm abuse**: People use algorithms for analysis, decision-making, coordination, and other activities, but their use purpose, use method, use range, etc., have deviations and cause adverse effects

Recognize and understand the potential AI ethical issues or risks

➢ **Long-term and indirect** ethical risks

- **Employment**: More and more work can be completed by some AI products.
- **Ownership**: A series of debates on ownership will follow, such as whether the AI agent should be considered as "legal subject," whether AI products have property rights (copyrights or patent rights).
- **Competition**: When companies, organizations or individuals use AI algorithms, they should follow competitive ethics and not go beyond legal boundaries.
- **Responsibility**: Who is responsible for bad consequences by AI.

Recognize and understand the potential AI ethical issues or risks

Four different categorizations of AI ethical issues

4. Categorization Based on the Deployment of AI

➢ Impact on **society**:

- **The labor market**: AI technologies may put current job classes at risk, eliminate positions, cause mass unemployment in many job sectors.

- **Inequality**: Revenues will be split across fewer people and individuals with ownership in AI-driven companies will receive disproportionate benefits, which indeed increase social inequalities.

- **Privacy, human rights, and dignity**: For instance, if AI can be used to determine people's political beliefs, then individuals may be vulnerable to manipulation. Bias: Human bias, such as gender prejudice and racism bias, may be inherited by AI

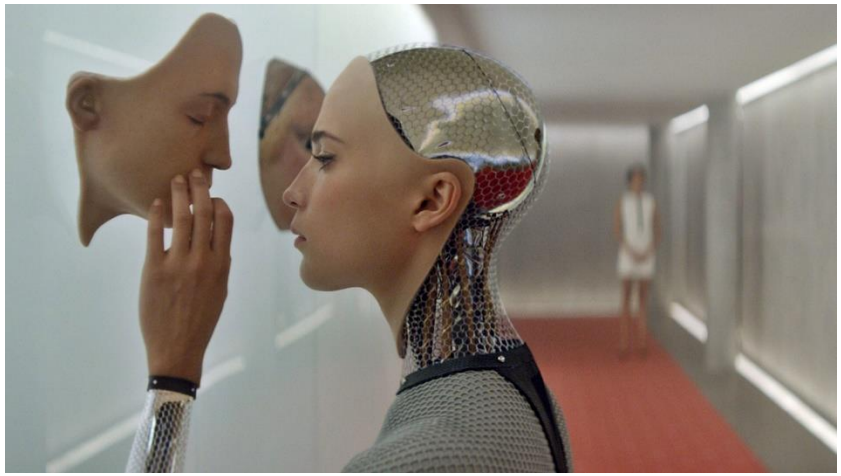
- **Democracy**: AI corporations could pose undue influence over the government. AI may damage democracy by affecting political elections

Recognize and understand the potential AI ethical issues or risks

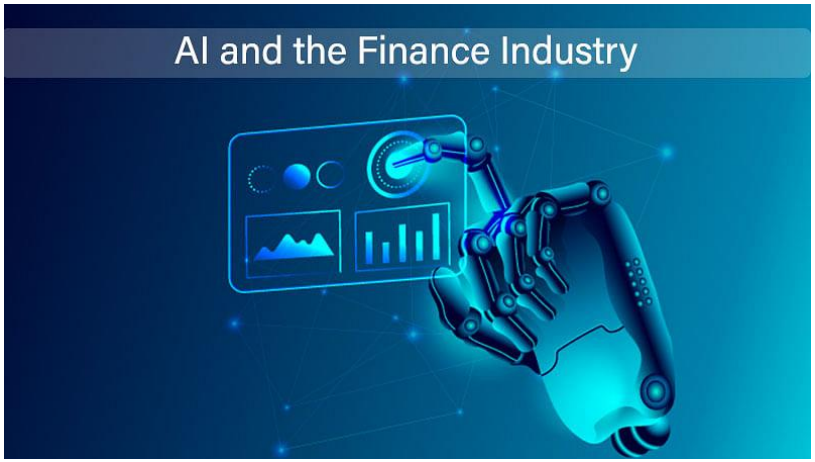
➢ Impact on **human psychology**

- **Relationships**: we will frequently interact with machines or AI products as if they are humans. This will have impacts on real human relationships and thus bring some ethical issues.

- **Personhood**: Whether AI system should be endowed with "personhood" and moral or legal agency rights.



Recognize and understand the potential AI ethical issues or risks



➢ Impact on **the financial system**

- The AI-based automatic trading agents may also **be used maliciously** to destabilize the markets or harm innocent parties in other ways.
- **Even if they are not intended to be malicious,** the autonomy and flexibility of algorithmic trading strategies, including the increasing use of ML techniques, make it difficult for people to predict how they will perform in unexpected situations

Recognize and understand the potential AI ethical issues or risks

➢ Impact on the **legal system**

- **Criminal law**: If AI products or robots are shown to have sufficient consciousness or awareness, then they may be the direct perpetrators of criminal offenses or responsible for negligent crimes.

  --- If we admit that AI products have their own mind, human-like free will, autonomy, or moral sense, then our criminal law and even the entire legal system will have to be revised.

- **Tort law**: Tort law covers situations such as one person's behavior case injury, suffering, unfair loss, or harm to another person.

  --- When an accident involving self-driving car(s) occurs, there are two legal areas that are relevant—negligence and product liability.

Recognize and understand the potential AI ethical issues or risks

➢ Impact on the **environment and the planet**

- **Use of natural resources**: The development and application of AI will increase the demand for many natural resources, such as rare earth metals like nickel, cobalt, graphite, and so on

- **Pollution and waste**: The increase in production and consumption of AI technological devices such as robots will exacerbate pollution and waste, such as the accumulation of heavy metals and toxic materials in the environment

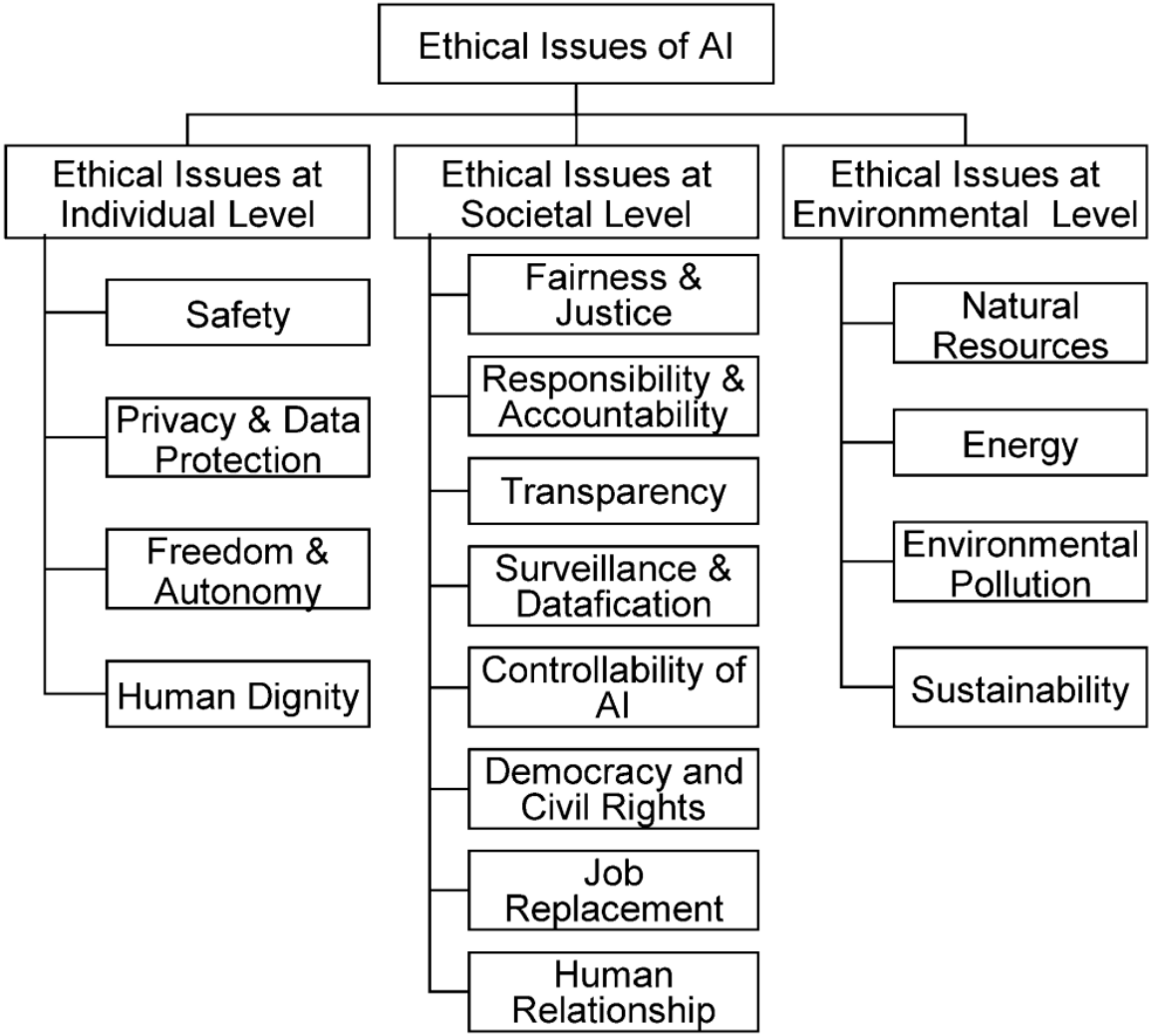
- **Energy concerns**: AI involves training ML models on a huge amount of data, which usually consumes large amounts of energy.

Recognize and understand the potential AI ethical issues or risks

➢ Impact on **trust**

- **Fairness**: As more and more decisions are delegated to AI, we must ensure that these decisions are free from bias and discrimination
- **Transparency**: Transparency is important for building trust in AI since it should be a must to know why an AI system made a particular decision, especially if that decision caused undesirable consequences or harm.
- **Accountability**: In the event of damages, accountability is essential to establish a remedial mechanism so that victims can receive adequate compensation.
- **Control**: as the intelligence of AI increases to the point that it surpasses human abilities, AI may come to take control over our resources and outcompete our species, and even leading to human extinction

A summary of AI ethical issues [1]



[1] C. Huang, Z. Zhang, B. Mao and X. Yao, "An Overview of Artificial Intelligence Ethics," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 4, pp. 799-819, Aug. 2023.

Ethical guidelines and principles for AI

- The number of AI guidelines issued each year from 2015 to 2021 is counted and listed in the table.

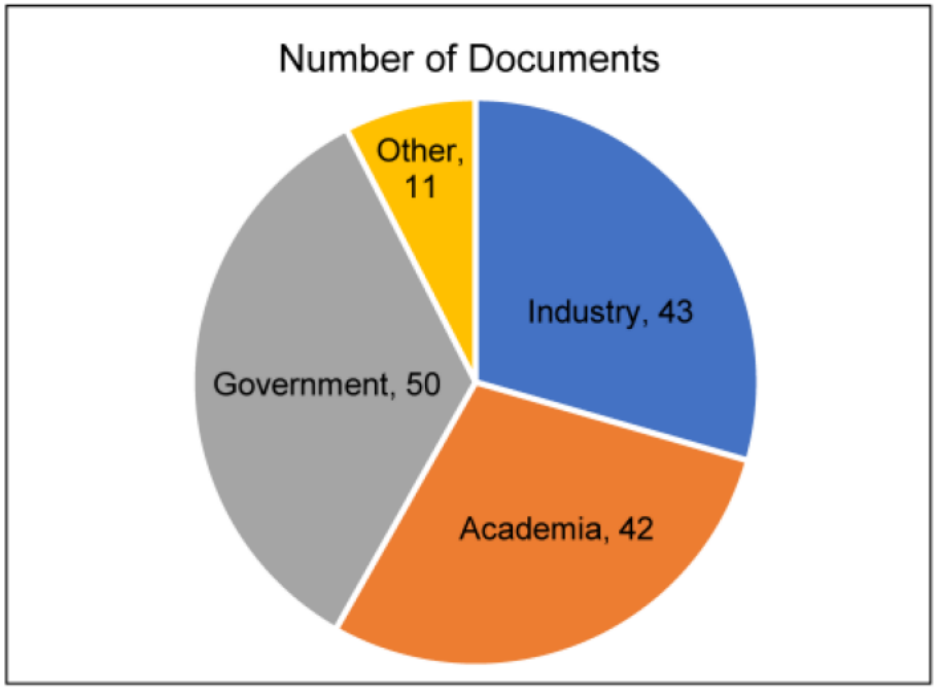
NUMBER OF DOCUMENTS ISSUED EACH YEAR FROM 2015 TO 2021

Year	2015	2016	2017	2018	2019	2020	2021
Number of Documents	2	7	25	53	31	24	4

NUMBER OF GUIDELINES ISSUED BY EACH COUNTRY OR REGION

Country	Australia	Canada	China	Denmark	EU	Finland	France
Number.	3	4	5	4	15	4	3
Country	Germany	Iceland	India	International	Ireland	Japan	N/A
Number	7	1	1	12	3	6	3
Country	Netherlands	Norway	Russia	Singapore	South Korea	Spain	Sweden
Number	4	1	1	3	3	2	1
Country	Switzerland	Turkey	UAE	UK	USA	Vatican	
Number	1	1	2	16	39	1	

Ethical guidelines and principles for AI



Percentage of guidelines released by different types of issuers.

- **Governments, companies, and academia all have shown strong concerns about AI ethics.**

Ethical guidelines and principles for AI

- The **11 ethical principles** identified in the existing AI guidelines

Freedom and Autonomy

Transparency

Fairness & Justice

Privacy

Responsibility and Accountability

Sustainability

Beneficence

Nonmaleficence

Dignity

Solidarity

Trust

IEEE guidelines of ethics

- IEEE is formed in 1963 as a merger of AIEE (American Institute of Electrical Engineers) and IRA (Institute of Radio Engineers)
- World's largest professional/technical organization for advancement of technology
- IEEE membership requires follow **IEEE guidelines/code of ethics**

IEEE

IEEE CODE OF ETHICS

WE, THE MEMBERS OF THE IEEE, in recognition of the importance of our technologies in affecting the quality of life throughout the world and in accepting a personal obligation to our profession, its members and the communities we serve, do hereby commit ourselves to the highest ethical and professional conduct and agree:

1. to accept responsibility in making decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;
2. to avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist;
3. to be honest and realistic in stating claims or estimates based on available data;
4. to reject bribery in all its forms;
5. to improve the understanding of technology, its appropriate application, and potential consequences;
6. to maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;
7. to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;
8. to treat fairly all persons regardless of such factors as race, religion, gender, disability, age, or national origin;
9. to avoid injuring others, their property, reputation, or employment by false or malicious action;
10. to assist colleagues and co-workers in their professional development and to support them in following this code of ethics.

IEEE guidelines of ethics for AI in law system

Ethical concerns for using AI in the law system	**VS.**	**Alignment with IEEE Guidelines of Ethics for AI**
Concerns about biased historical data perpetuating biases in the criminal justice system		Ethical AI development should strive for fairness and avoid bias
Lack of complete legal reasoning and individualized explanations in AI decisions		Importance of transparency and accountability in AI systems
Incomplete and edited nature of available data compromising training data quality		AI developers must ensure reliable, unbiased, and representative data

IEEE guidelines of ethics for AI in law system

Rapid changes in legislation and judicial interpretations challenging outdated AI systems	Importance of complying with legal standards and updating AI systems
AI's impact on individuals' lives and potential infringement on human rights	Prioritizing human well-being and respecting human rights in AI development
Gender-based disparities in sentencing perpetuated by AI systems	Ethical AI development requires identification and mitigation of biases
Privacy issues related to the "China Judgements Online" database	Respecting privacy rights in AI development

IEEE guidelines of ethics for AI in law system

Challenges in understanding and accurately applying complex legal reasoning	Designing AI systems to navigate legal complexities and provide just recommendations

❑ For Addressing Concerns

- Transparent and inclusive practices
- Regular bias audits and collaboration with legal and ethical experts
- Monitoring and updating AI systems to adapt to changes in laws and societal norms

. . .

Case Study: AI in Ride Sharing

Background: A tech company is developing a ride-sharing platform that aims to efficiently connect drivers and passengers in a busy urban environment to reduce their travel time by using AI algorithms.

What are the **Trade-offs** and **considerations**?

Case Study: AI in Ride Sharing

❑ Algorithmic Efficiency vs. Fairness

Trade-off: The AI algorithm is known for its efficiency in finding the shortest path between two points.

However, optimizing solely for efficiency might lead to situations where certain areas or users consistently experience longer wait times. How can the algorithm balance efficiency with the fair distribution of services?

❑ User Privacy vs. Personalization

Trade-off: To optimize routes effectively, the algorithm needs access to user location data.

However, this raises concerns about privacy. How can the platform personalize services based on user preferences and historical data while respecting and protecting user privacy?

Case Study: AI in Ride Sharing

❑ Cost Optimization vs. Driver Compensation

Trade-off: Minimizing travel times can be achieved by optimizing routes that might involve cost-cutting measures, such as reducing driver compensation or prioritizing certain trips over others.

How can the company balance the need for cost optimization with fair compensation for drivers?

❑ Real-Time Updates vs. User Experience

Trade-off: AI can dynamically adjust routes based on real-time traffic data, providing more accurate estimates.

However, constantly changing routes may lead to confusion and frustration for users. How can the system provide real-time updates without compromising the overall user experience?

Case Study: AI in Ride Sharing

❑ Algorithmic Complexity vs. Scalability

Trade-off: AI is computationally efficient for finding optimal routes, but as the number of users and requests increases, the complexity of the algorithm may impact system scalability.

How can the platform maintain algorithmic efficiency while ensuring scalability to handle a growing user base?

❑ Incentives for Pooling vs. Individual Trips

Trade-off: Encouraging users to share rides can contribute to reduced traffic and environmental impact.

However, some users may prefer the convenience of individual trips. How can the system incentivize ride-sharing without alienating users who prefer solo trips?

Case Study: AI Ethics in Ride Sharing

Discussion Points

- How should the algorithm **prioritize efficiency versus fairness** in distributing ride-sharing services?
- What measures can be implemented to **address privacy concerns while still personalizing the user experience**?
- How can the company balance the need for **cost optimization with ensuring fair compensation for drivers**?
- What strategies can be employed to provide **real-time updates without compromising the user experience**?
- How might the company address **algorithmic complexity to ensure scalability as the user base grows**?

Case Study: AI Ethics in Ride Sharing

Discussion

Potential Arguments	VS.	Resolutions
• Efficiency Priority Some argue that prioritizing efficiency benefits all users by minimizing wait times and travel durations. • Fairness Priority Others argue that a truly equitable system ensures fair access, even if it means slightly longer wait times for some users.		• Implement dynamic prioritization considering factors like user location, historical wait times, and demand fluctuations.
• Personalization Benefits Personalized services enhance the user experience and satisfaction. • Privacy Concerns Users have the right to control their personal data, and excessive data collection may lead to privacy breaches.		• Implement strict data anonymization and aggregation protocols. • Allow users to opt in or out of data sharing for personalization, emphasizing transparency and user control.

Case Study: AI Ethics in Ride Sharing

• Cost Efficiency Reducing costs can contribute to lower prices for users and increased profitability for the company. • Fair Compensation Drivers should receive fair wages for their services to ensure a sustainable and ethical business model.	• Establish a transparent and fair compensation structure for drivers. • Consider implementing cost-cutting measures that don't compromise driver earnings, such as optimizing operational processes.
• Accuracy Priority Real-time updates based on current traffic conditions contribute to accuracy and efficiency. • User Experience Priority Constantly changing routes may lead to confusion and frustration among users.	• Provide real-time updates with clear communication. • Allow users to choose between dynamic updates and stable routes, ensuring transparency and choice.

Case Study: AI Ethics in Ride Sharing

• Efficiency Priority Efficient algorithms provide better service, especially during peak times. • Scalability Priority The system must accommodate growth without sacrificing performance or causing disruptions.	• Continuously optimize algorithms for efficiency and invest in scalable infrastructure. • Conduct regular performance assessments and updates to ensure both efficiency and scalability.
• Environmental Benefits Encouraging ride-sharing contributes to reduced traffic and environmental impact. • User Preference Users have the right to choose individual trips for convenience or privacy reasons.	• Implement dynamic pricing, loyalty programs, or environmental incentives for ride-sharing. • Respect user preferences and provide clear communication about the benefits of pooling.

References

- C. Huang, Z. Zhang, B. Mao and X. Yao, "An Overview of Artificial Intelligence Ethics," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 4, pp. 799-819, Aug. 2023.
- A. Morby, “Tesla driver killed in first fatal crash using autopilot,” 2016. Accessed: Feb. 10, 2022. [Online]. Available: https://www.dezeen.com/2016/07/01/tesla-driver-killed-car-crashnews-driverless-car-autopilot/
- S. McGregor, Ed., “Incident number 6,” in AI Incident Database, 2016. [Online]. Available: https://incidentdatabase.ai/cite/6
- R. V. Yampolskiy, “Predicting future AI failures from historic examples,” Foresight, vol. 21, no. 1, pp. 138–152, 2019.
- C. Stupp, “Fraudsters used AI to mimic CEO’s voice in unusual cybercrime case: Scams using artificial intelligence are a new challenge for companies,” 2019. Accessed: Feb. 10, 2022. [Online]. Available: https://www.wsj.com/articles/fraudsters-use-ai-tomimic-ceos-voice-in-unusual-cybercrime-case-11567157402