Digital Ethics and Data Privacy

Topic 2: Artificial Intelligence Ethics and Governance

Case Scenario: Ethical Dilemma in AI

- Imagine you're the lead data scientist in a tech firm developing an AI-driven facial recognition system for law enforcement.
- During a meeting, your team discovers that the AI system is showing an accuracy discrepancy. It is less accurate in recognizing faces of people with darker skin tones, potentially leading to wrongful identifications.
- Fixing this issue would require extensive retraining of the model, which would push back the project timeline and exceed the budget.

1

Continue development without changes

No AI system is perfect

Law enforcement to use the system responsibly

2

Pause the development and address the issue

Flaw can lead to harmful consequences

AI system should not disadvantage any group

3

Release current version and perform continuous improvements

Balance approach to handle constraints with need for fair AI system

Case Scenario: Explainability, Transparency, and Fairness

- You are a machine learning engineer in a financial technology (fintech) company that uses AI models to evaluate and approve loan applications.
- The AI system used has been delivering excellent results in terms of risk prediction and minimizing defaults. However, it operates as a "black box" model, and its decision-making process isn't easily explainable.
- This leads to complaints from rejected applicants who want to understand the reasons behind their loan denial.

1. Stick with the less explainable AI system

   Accuracy and success is more important

2. Advocate for new more explainable AI system

   Trust-building and fairness more important

3. Propose a hybrid approach

   Balance approach - efficiency and risk minimization

Case Scenario: Human-Centric Approach in AI

- Imagine you're the project lead at a tech company developing an AI-based productivity tool that uses machine learning algorithms to automate and optimize tasks for employees.
- During the testing phase, you realize that the tool, while successful in automating tasks, is also recommending job cuts in certain areas, threatening the job security of a significant number of employees.

1

Proceed with the tool

Streamline processes and improve efficiency

Reskill or upskill affected employees

2

Modify the tool to focus more on assisting employees

Enhance their productivity and job satisfaction

3

Implement the tool with a transition plan with new roles training

Balance approach – AI's efficiency benefits + human impact mitigation

Case Scenario: Death Predictor

- The death predictor - a highly complex AI algorithm that scans a patient's electronic health record and calculates that person's chance of dying within a specified period of time.

- It's a deep learning algorithm that evaluates more than 12,000 pieces of information from a patient's medical records within 24 hours of a patient being admitted to SGH. It looks at everything from age, gender, race, to disease classifications, prescription codes, to doctor's notes. And then it generates a mortality prediction.

Role Sheet

Stakeholders	Key Questions to Consider
Patients: This group represents the individuals who are directly affected by the death predictor system.	**Role Sheet Questions**: • How would you feel about a prediction system determining your chance of death within a specified time? • What are your thoughts on your doctor initiating end-of-life care based on this prediction? • How might knowing this prediction impact your mental health and relationships? • How would you react if the prediction was wrong?
Doctors and Medical Staff: This group will interact with the system and communicate its results to the patients.	**Role Sheet Questions**: • How would this system impact your relationship with patients? • How would this affect your decision-making and approach towards patient care? • What would you do if you disagreed with the prediction?

Role Sheet

Stakeholders	Key Questions to Consider
Hospital Administration: This group will make the decision to implement the system and manage its consequences.	**Role Sheet Questions**: • How does implementing this system align with the hospital's mission and ethical guidelines? • What are the potential legal implications if the system makes incorrect predictions? • How would you manage potential backlash or controversy related to the system?
Family Members of Patients: They will be emotionally affected by the system's predictions and involved in end-of-life decisions.	**Role Sheet Questions**: • How would you feel about a system predicting the likelihood of your loved one's death? • What impact could this have on your relationship with the patient and the medical staff? • How could this prediction affect end-of-life decision-making for your loved one?

Role Sheet

Stakeholders	Key Questions to Consider
AI Developers and Researchers: They develop the system and have a responsibility to ensure it is ethical, fair, and accurate.	**Role Sheet Questions**: • How can you ensure that the AI system is unbiased and accurate? • How can you handle the ethical implications of making death predictions? • How can the system be improved based on user feedback and errors? • How can you ensure that the system respects patient privacy and autonomy?

Read, Do and Watch

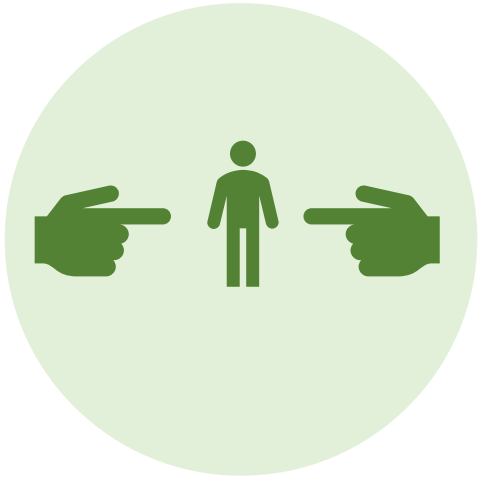
Topic Objectives

- Identify what is AI and ethical issues of AI applications.
- Outline the three generations of AI and classification of AI types.
- Learn and apply the five pillars of trustworthy AI to AI applications.
- Understand the need for governance in AI.
- Evaluate and apply an AI Ethics Governance Framework to guide responsible implementation of AI systems.
- Examine case studies of ethical issues in AI, such as bias and fairness.

AI and AI Ethics

- Artificial intelligence (AI) is a term to describe a branch of computer science that is dedicated to creating intelligent machines that can do work and react like humans.
- AI Ethics is a set of guidelines that advise on the design and outcomes of AI-enabled systems

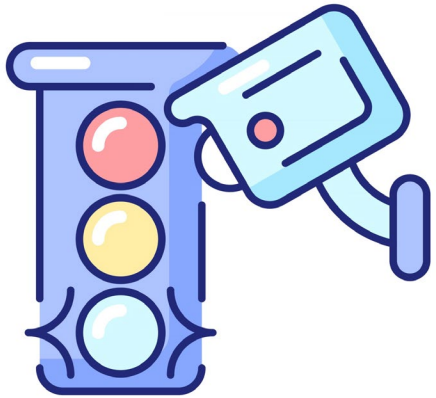
Privacy and Surveillance



Bias and Discrimination

Human Judgment

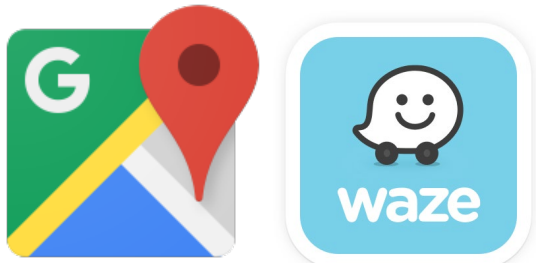
AI Ethics: Privacy and Surveillance

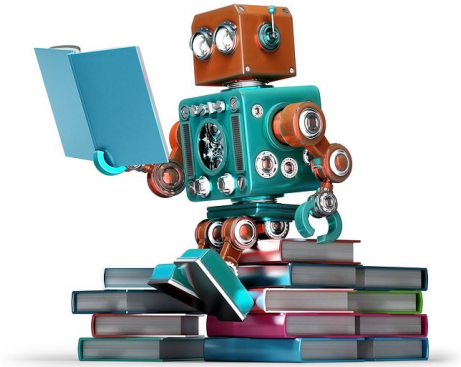


A security guard monitors people through his augmented reality eyewear equipped with an infrared temperature detector in Hangzhou, China, March 24, 2020. © 2020 AP Photo

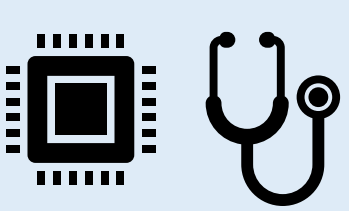
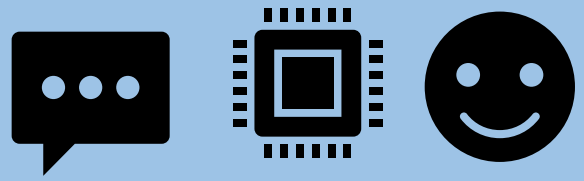
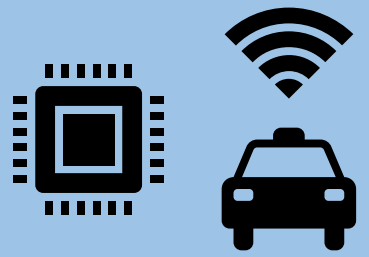
Privacy and Surveillance

AI At Work and School

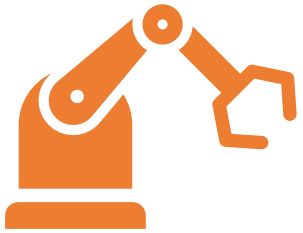




3 Generations of AI

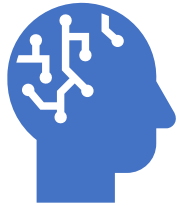
Expert Systems	Speech/Facial Recognition	Self-driving Cars
		
First Generation	Second Generation	Third Generation
- Knowledge - Algorithms - Computing Power	- Data - Algorithms - Computing Power	- Knowledge - Data - Algorithms - Computing Power

4 Types of Artificial Intelligence

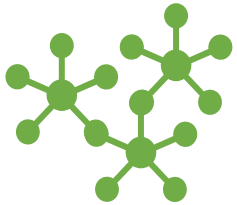


Reactive Intelligence

Limited Memory Intelligence

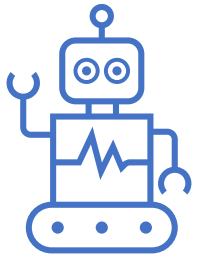


Theory of Mind

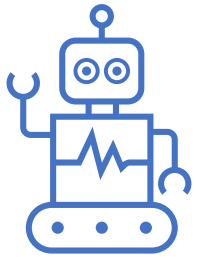


Self-Aware AI

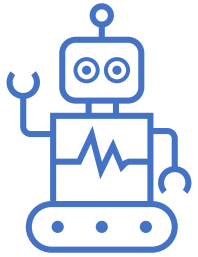
3 Laws of Robot



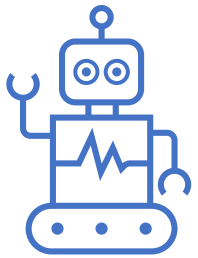
1st Law: A robot may not injure a human or, through inaction, allow a human being to come to harm.



2nd Law: A robot must obey the orders given to it by a human being except where such orders would conflict with the First Law.



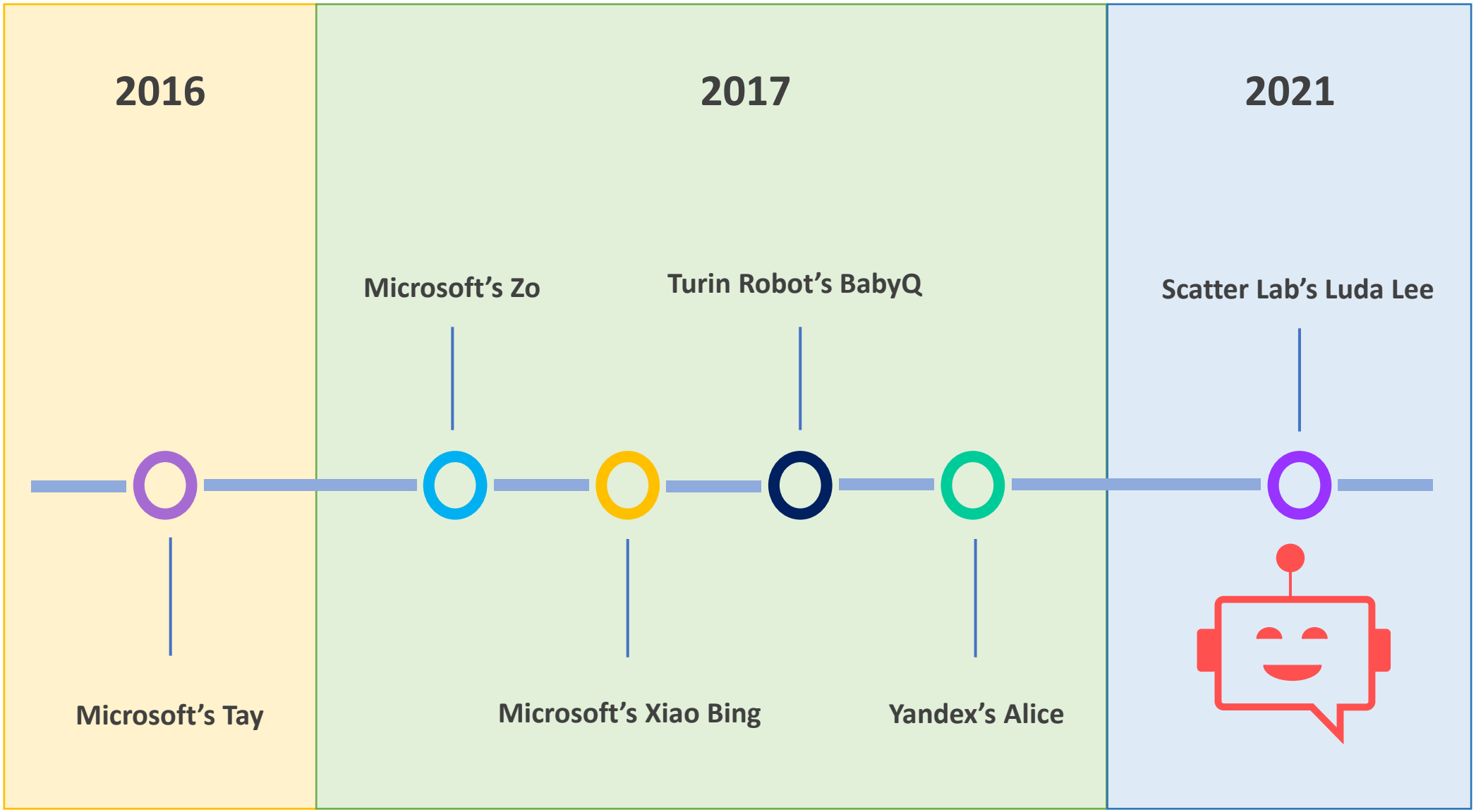
3rd Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.



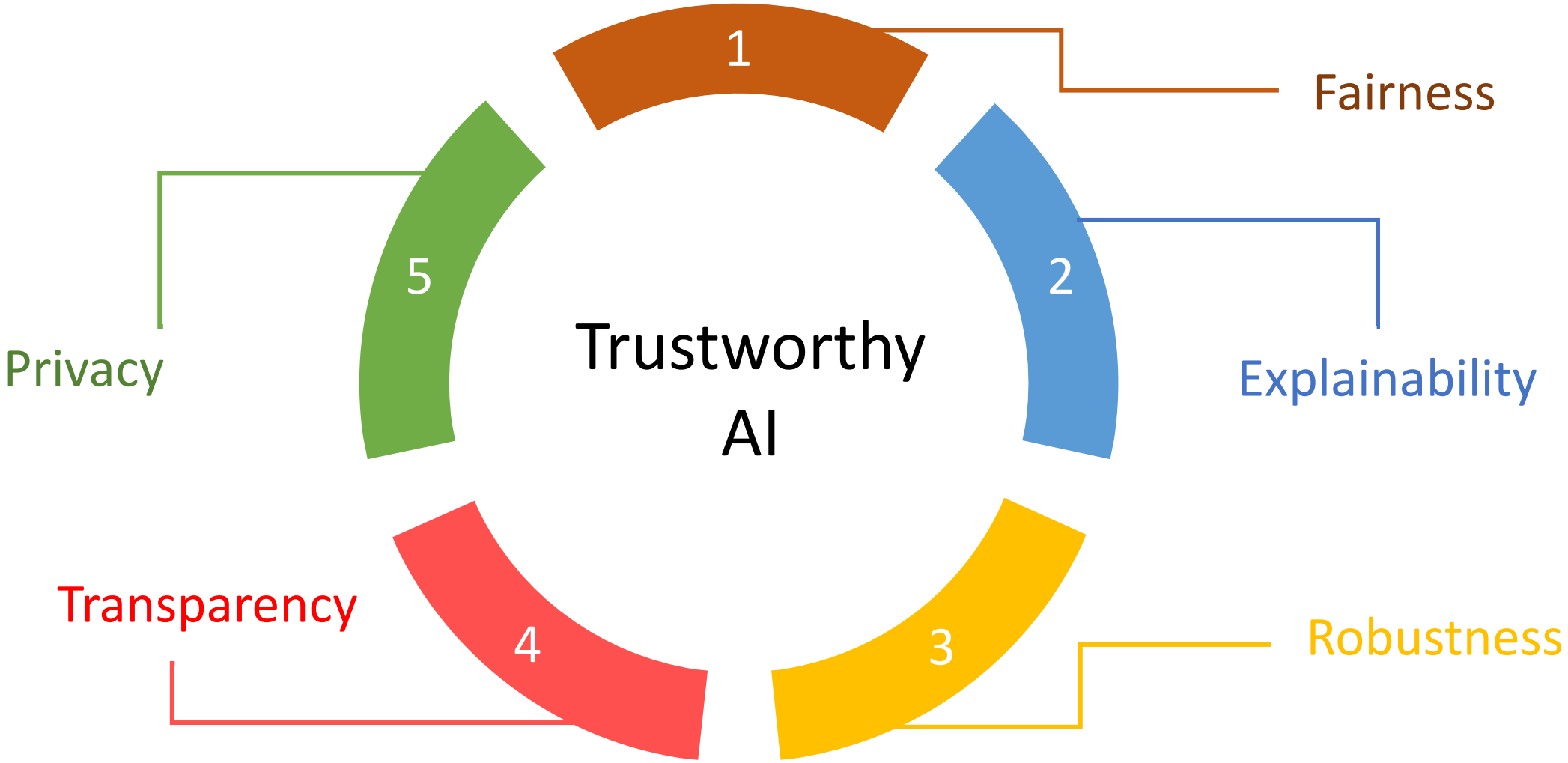
0th Law:

A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

Epic Chatbot Failures in Social Engagement

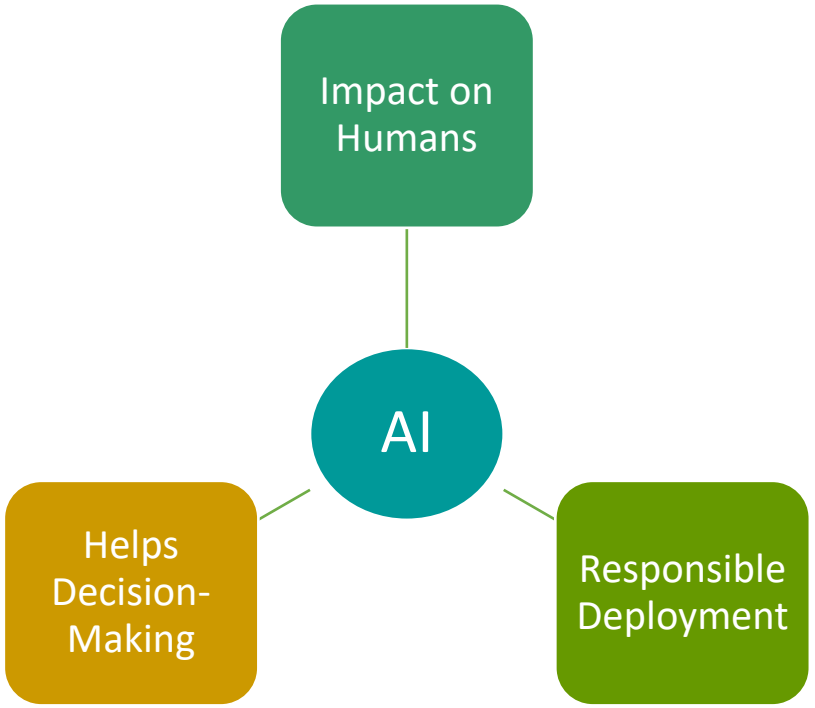


Five Pillars of Trustworthy AI



Source https://www.ibm.com/artificial-intelligence/ethics

Why Care for AI Ethics and Governance?



Source: AI Ethics and Governance BoK and Toolkit Developed by SCS and IMDA

AI Governance

- AI governance is the idea that there should be a legal framework for ensuring that AI and machine learning technologies are well researched and developed with the goal of helping humanity navigate the adoption of AI systems fairly.

Body of Knowledge (BoK)

- The Body of Knowledge provides guidance to business leaders and professionals on how to deploy AI responsibly.

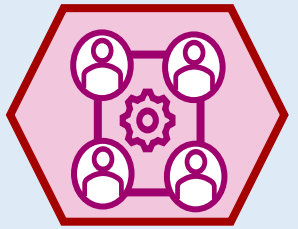
Two Principles and Four Pillars For AI Adoption

Two guiding principles	1. Decisions made by AI should be "explainable, transparent and fair" 2. AI systems should be human-centric

Internal Governance

Operations Management

Human-Centricity



Stakeholder Communications

Proliferation of National AI Strategies

NUS National University of Singapore | School of Computing

Next Generation AI Development Plan for China	Artificial Intelligence for Europe	Executive Order on the American AI Initiative
"…to build China's first-mover advantage in the development of AI…"	"…a leading role in setting the global gold standard…"	...maintaining American leadership in AI..."

"Whoever becomes the leader in this sphere will become the ruler of the world"

- Putin

Singapore's National AI Strategy

Smart Nation SINGAPORE

Vision

By 2030, we see Singapore as a leader in developing and deploying scalable, impactful artificial intelligence (AI) solutions, in key sectors of high value and relevance to our citizens and businesses.

Objectives

Global Hub for AI Solutions

Govern and Manage AI Impact

Generate Economic Value & Improve Lives

Singapore's National AI Projects

NUS National University of Singapore | School of Computing

Transport and Logistics

INTELLIGENT FREIGHT PLANNING

Smart Cities and Estates

SEAMLESS & EFFICIENT MUNICIPAL SERVICES

Healthcare

CHRONIC DISEASE PREDICTION & MANAGEMENT

Education

PERSONALISED EDUCATION THROUGH ADAPTIVE LEARNING & ASSESSMENT

Safety and Security

BORDER CLEARANCE OPERATIONS

Further Reading

1. China's Xinjiang surveillance

https://www.engadget.com/2018-02-22-china-xinjiang-surveillance-tech-spread.html?_fsig=Z7sJRecikk0HcHN2kOvxVQ--%7EA

2. Generative AI at school, work and the hospital

https://theconversation.com/generative-ai-at-school-work-and-the-hospital-the-risks-and-rewards-laid-bare-227967

3. 10 ways artificial intelligence is changing the workplace

https://www.businessinsider.com/ai-transforming-the-workplace-examples-2023-7

4. The brief history of artificial intelligence

https://ourworldindata.org/brief-history-of-ai

5. Understanding the Four Types of Artificial Intelligence

https://iabac.org/blog/understanding-the-four-types-of-artificial-intelligence

6. AI ETHICS AND GOVERNANCE BODY OF KNOWLEDGE

https://www.scs.org.sg/ai-ethics-bok

Q & A


NUS
National University
of Singapore
School of
Computing