CSN - 391 Technical Communication

ETHICS IN AI
(Artificial Intelligence)

Dhruv Agrawal (22114030)
Ischit Kumar(22114038)
Manharlal(22114019)

Himanshu Raheja(22323023)
Komal(22113078)

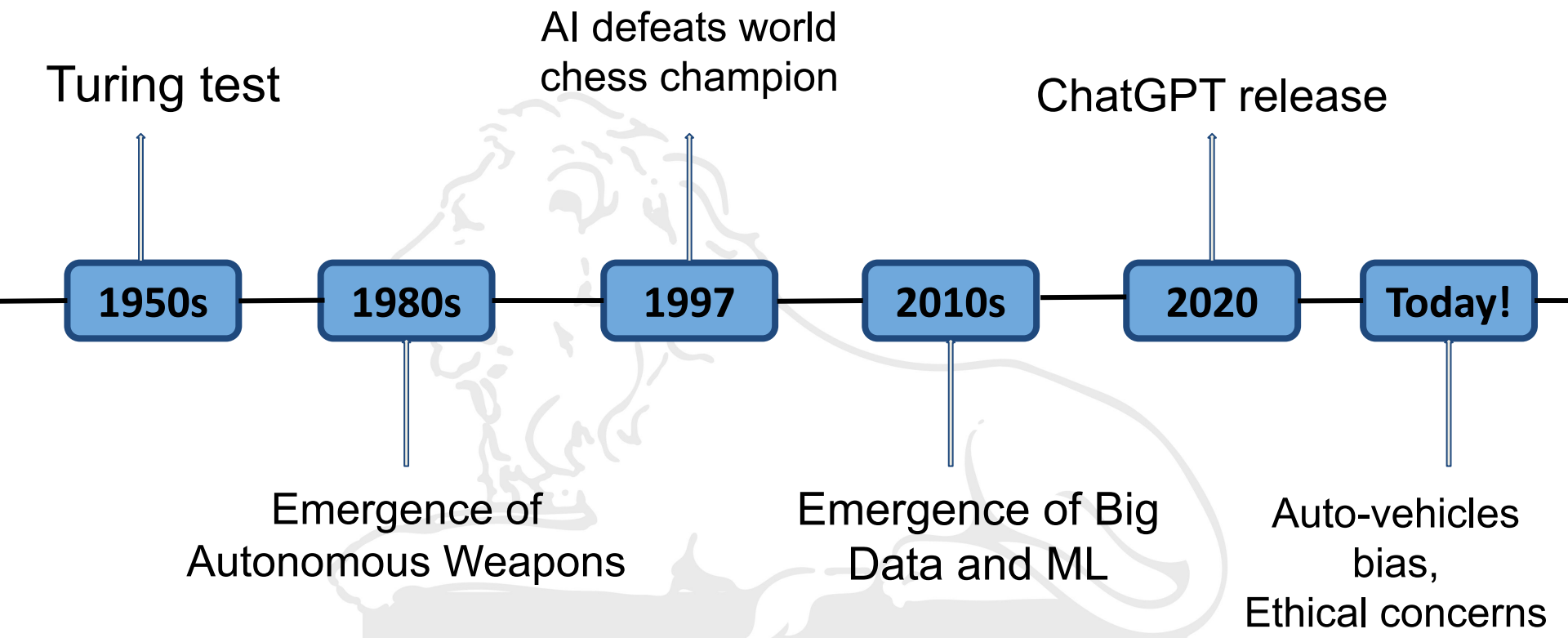
Contents

How did “Ethics in AI” come into the picture?

"With artificial intelligence, we are summoning the demon."

- Elon Musk

How did “Ethics in AI” come into the picture?



Definition of AI Ethics

What is AI Ethics?

AI Ethics refers to the guidelines and principles that govern the development and deployment of artificial intelligence to ensure it aligns with human values and ethical standards.

Why is Ethics Important in AI

- AI systems are increasingly influential in decision-making processes across various sectors.
- Ethical considerations ensure AI benefits society while minimizing harm and unintended consequences.

Historical Context and Evolution

Early Ethical Considerations:

Initially, AI research focused on technical aspects, but ethical concerns grew as AI systems began affecting human decisions.

Evolution of AI Ethics:

Ethical discussions have shifted from theory to practical frameworks, addressing issues like fairness, transparency, and accountability.

Key Milestones:

1942: Asimov’s "Three Laws of Robotics."
2016: Formation of the Partnership on AI to tackle societal impacts

Core Ethical Principles in AI

Core ethical principles in AI are fundamental guidelines that ensure the responsible design, deployment, and operation of AI systems, balancing innovation with societal impact.

1

Transparency

Refers to making the processes, decisions, and outcomes of AI systems understandable and traceable.

2

Fairness

Involves ensuring that AI systems do not discriminate or exhibit bias towards individuals or groups.

3

Accountability

Incorporates the responsibility and liability for the actions and decisions made by AI systems.

4

Privacy

Focuses on safeguarding confidential and sensitive data handled by AI systems to protect user privacy.

5

Security

Addresses the protection of AI systems from unauthorized access, cyber threats, and malicious activities.

AI in specific sectors

Use of AI in real life

1. Improved efficiency and productivity .
2. Improved accuracy .
3. Personalisation of choice .
4. Solving complex problems.
5. Reducing human error.

Using AI in medical field

- **Medical diagnosis**
  AI enhances accuracy and speed in medical diagnosis.

- **Personalized Medicine**
  AI customizes treatments to individual needs, enhancing outcomes and precision.

- **Drug discovery**

Using AI in education

- **Personalised learning**
  Customizes educational content

- **Tutoring Systems**
  AI-powered tutoring provides additional support while maintaining fairness and avoiding biases.

- **Administrative Efficiency**
  AI automates tasks, streamlining administration and reducing educator workloads.

Using AI in financing

- **Fraud Detection**
  AI detects fraudulent transactions, enhancing security and preventing financial crimes.

- **Risk Assessment**
  AI evaluates credit and investment risks, ensuring fair and accurate assessments.

- **Customer Service**
  AI chatbots provide efficient support to the consumers .

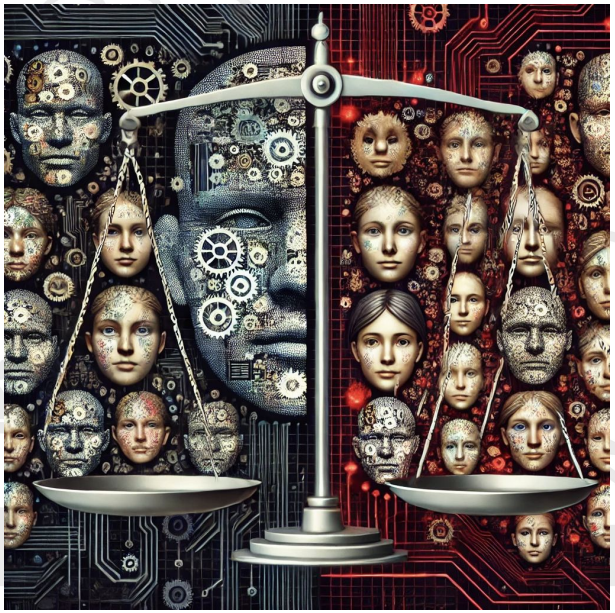
Ethical Issues in AI

Ethical Issues in AI

Bias and Fairness

Algorithm Bias

AI systems can inherit biases from training data, leading to unfair outcomes



Case in Point

In 2018, Amazon scrapped an AI tool that favored male resumes and penalized terms like "women's."

Consequences of Bias

AI systems can inherit biases from training data, leading to unfair outcomes.

Mitigation Strategies

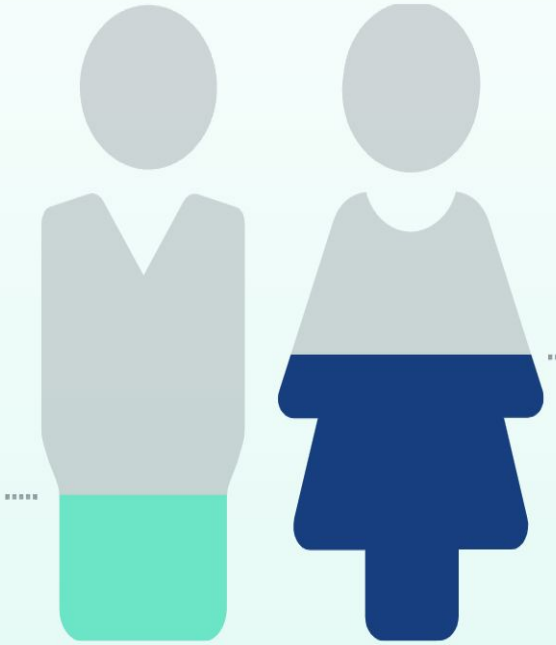
AI systems can inherit biases from training data, leading to unfair outcomes.

Ethical Issues in AI

BIAS IMPACT ANALYSIS

Impact of Bias in Algorithms

Language models perpetuating covert racism against speakers of African American English (AAE).



Job Assignments

23%

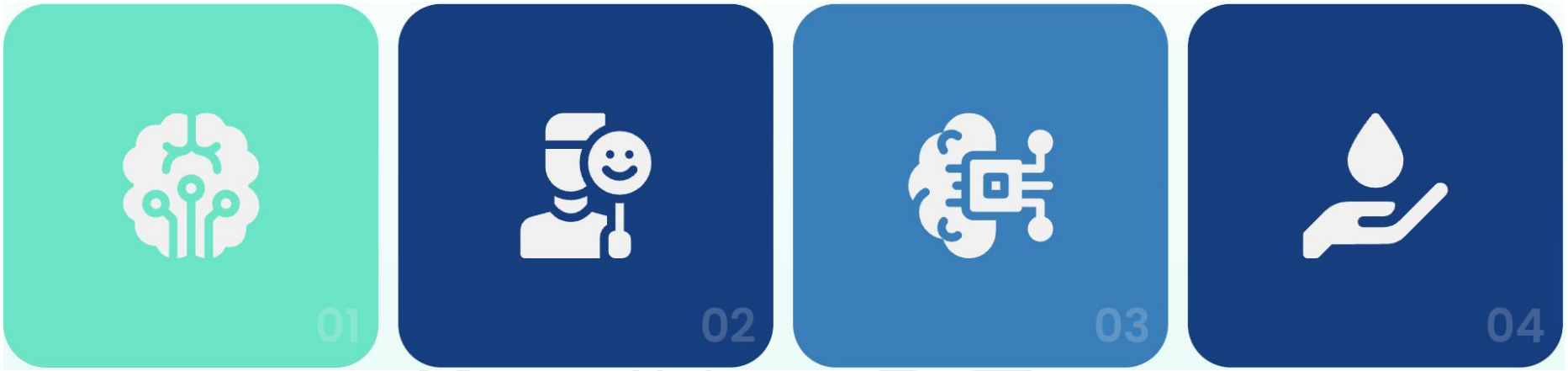
45%

Likelihood of Death Penalty

Hofmann et al., 2024

Ethical Issues in AI

Privacy & Data Protection



Data Collection

AI systems often require large amounts of personal data, raising concerns about how this data is gathered and used.

Lack of Consent

Users may not be fully aware of or have not consented to how their data is being used by AI systems.

Case in Point

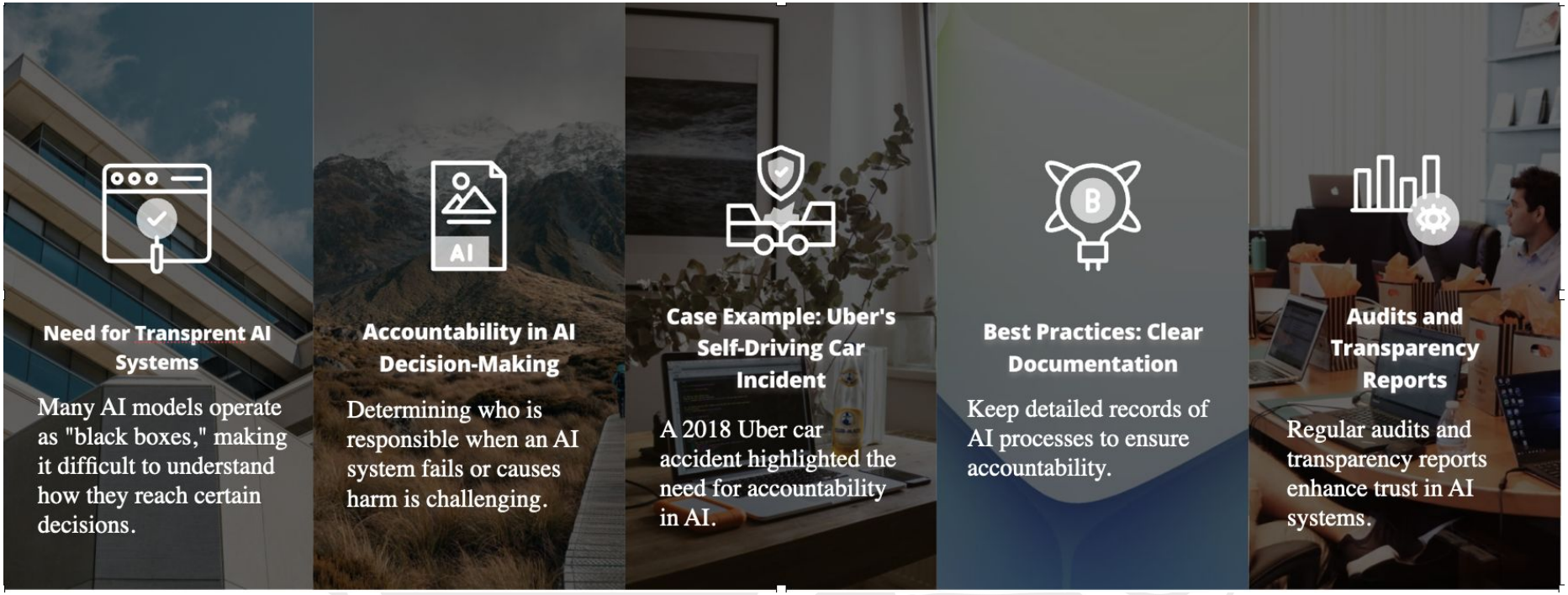
China's use of facial recognition to monitor the Uyghur population in 2019 raised global concerns about privacy violations.

Mitigation Strategy

Limit data collection to only what's necessary, and ensure transparency about how data is used to protect user privacy.

Ethical Issues in AI

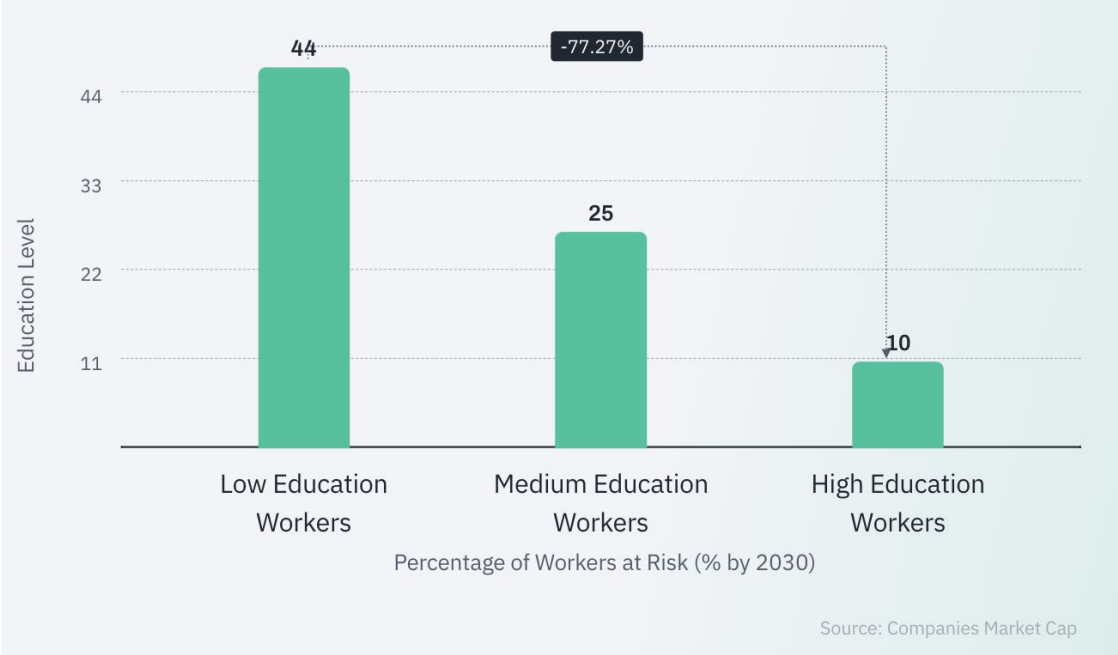
Transparency and Accountability



Ethical Issues in AI

Job Displacement

AI and automation are disrupting job markets, potentially leading to significant unemployment.



In 2023, companies like Meta used AI for content moderation, leading to significant layoffs as AI replaced human moderators.

Ethical Issues in AI

Autonomous System & Safety

Lack of Human Intervention

Autonomous systems pose risks due to the absence of immediate human control or decision-making, potentially leading to unforeseen consequences.

Safety issues in self driving cars

Self-driving cars face ethical dilemmas in scenarios where decisions must be made, such as choosing between passenger safety and pedestrian well-being.

Concerns in drone operations

Drones raise privacy and safety concerns, especially in crowded areas, highlighting the need for strict regulation

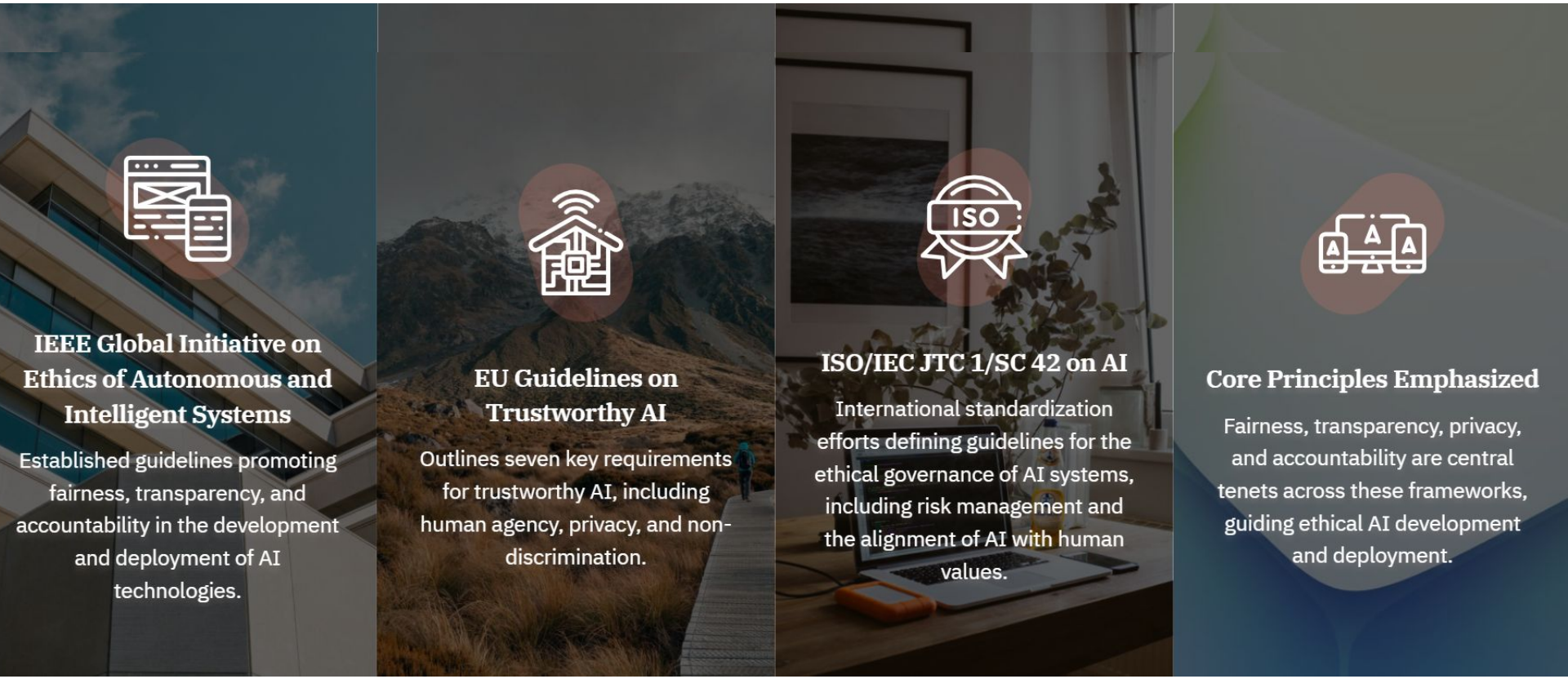
Unforeseen Consequences

The complex interactions of autonomous systems can result in unexpected outcomes that may not have been accounted for during development

"How can we ensure that AI systems are not only powerful but also responsible and fair in their impact on society?"

Current AI Frameworks

Exploring Key Ethical Frameworks Shaping AI Development



Ethical frameworks are attempts to build consensus around **values and norms** that can be adopted by a community – whether that's a group of individuals, citizens, governments, businesses within the data sector or other stakeholders.

Ethical black box Initiatives including the **UNI Global Union** and **IEEE** suggests:

- AI systems with an '**ethical black box**': a device that can record information about said system to ensure its accountability and its transparency,
- It should also includes clear data on the ethical consideration built into the system from the beginning.

The idea is similar to the black box in airplanes, which records flight data to help understand events in case of an incident.

Documentation

Transparency

Accountability

Considers Ethics

'through every digital transaction (explicit or observed) humans are generating a unique digital shadow of their physical self' (IEEE, 2017)

European Union

PII (Personally Identifiable Information) has been established as the asset of the individual (**by Regulation (EU) 2016/679 in Europe**, for example). One of the most significant frameworks is the General Data Protection Regulation, or GDPR, which has set a global standard for data privacy and protection.

Systems must ask for explicit consent at the time data are collected and used, in order to protect individual.

Autonomy

Dignity

Right to consent

NITI Aayog (#AIforAll)

Part 1 : Principles for Responsible AI

System Considerations:

- Decision-making principles for AI.
- Rightful inclusion of beneficiaries.
- Accountability of AI decisions.

Societal Considerations:

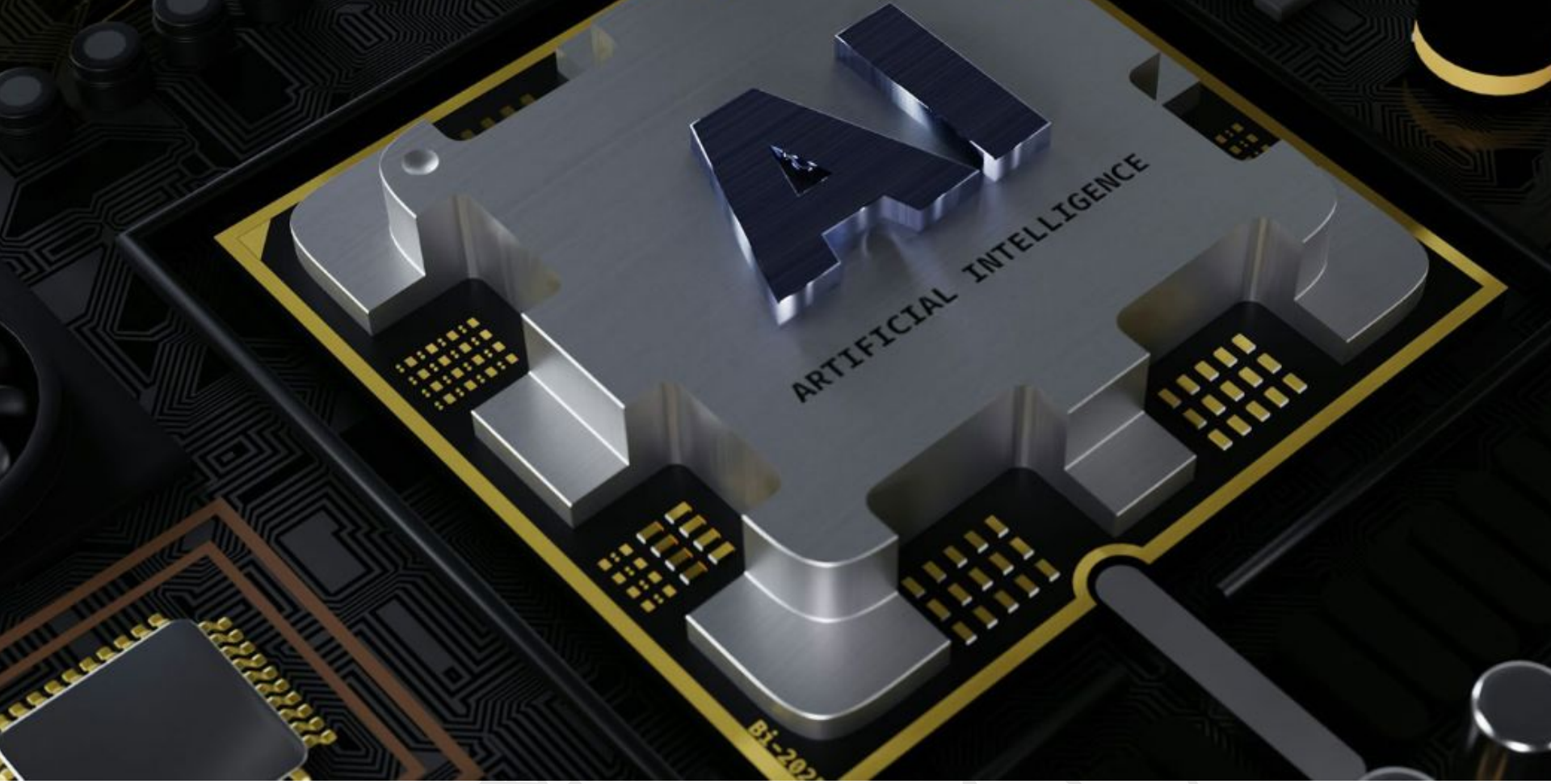
- Impact of automation on job creation.
- Effects on employment

Part 2 : Operationalizing Principles for Responsible AI

Focus Areas:

- Regulatory and policy interventions.
- Capacity building for responsible AI.
- Incentivizing ethics by design.
- Creating frameworks for compliance with AI standards.

India is a member of the *Global Partnership on Artificial Intelligence (GPAI)*. GPAI's Experts produce deliverables that can be integrated into Members' national strategies to ensure the inclusive and sustainable development of AI. Under the 2023 themes of climate change, global health and societal resilience, Experts worked to ensure that AI is used responsibly to address current challenges around the world.



CASE STUDY of Ethical AI Implementation

Successful implementation examples: Google | Integration of ethical principles into AI practices.

Core Elements of **SAIF**(Secure AI Framework)

Automate Defenses to keep pace with existing and new threats

Expand strong security foundations to the AI ecosystem

Contextualise AI system risks in surrounding business processes

Core Elements of **SAIF**(Secure AI Framework)

Adapt controls to adjust mitigations and create faster feedback loops for AI deployment

Harmonise platform level controls to ensure consistent security across the organisation

Extend detection and response to bring AI into an organisation's threat universe

The Role of REGULATION in AI

Are ethical guidelines sufficient to ensure that such principles are followed?

1. General Data Protection Regulation (GDPR) - Europe
2. California Consumer Privacy Act (CCPA) - USA
3. Digital Personal Data Protection Act (DPDPA) - India
4. Artificial Intelligence and Data Protection Bill - India (Proposed)

THE FUTURE OF AI ETHICS

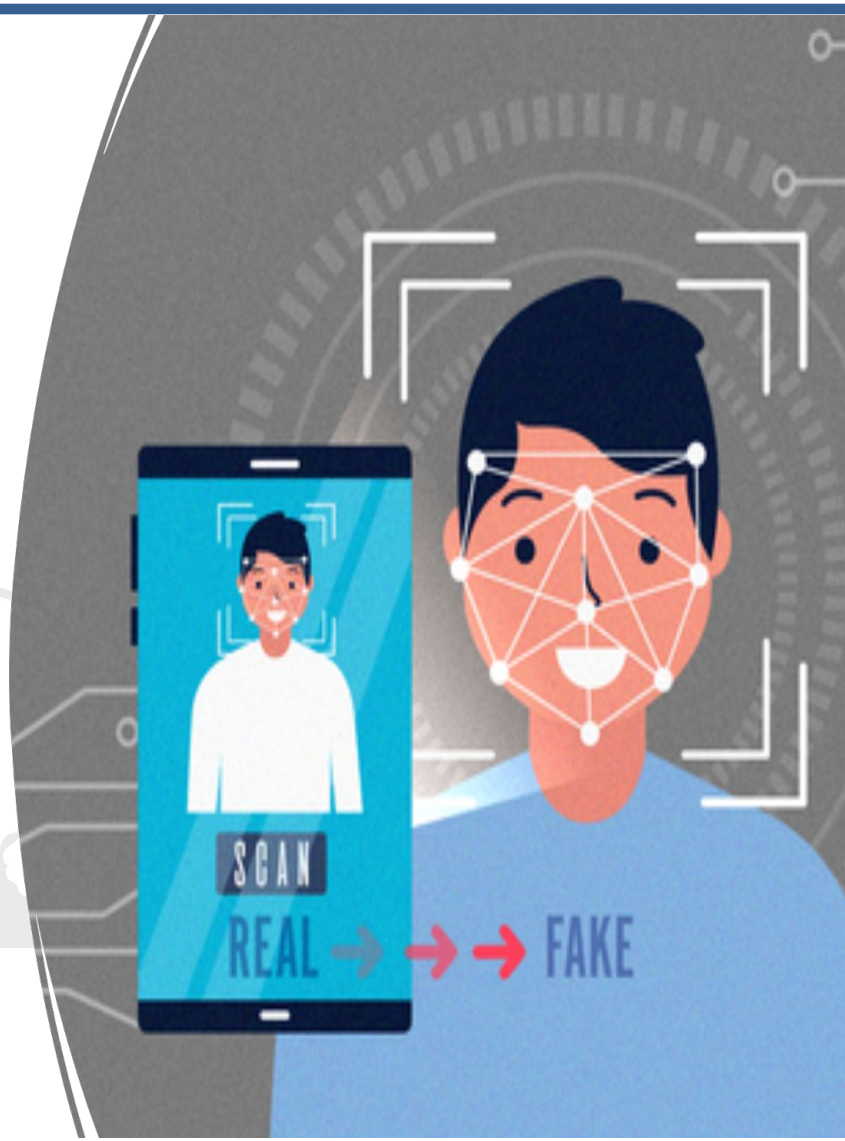
Exploring ethical considerations in AI for a sustainable future.

AI in Warfare and Autonomous Weapons

- **Autonomous Weapon System(AWS):** Development of AI – powered weapon system
- **Moral Responsibility:** Determining who is accountable
- **International Law and Human Rights:** Use of AI in warfare complies with international humanitarian law

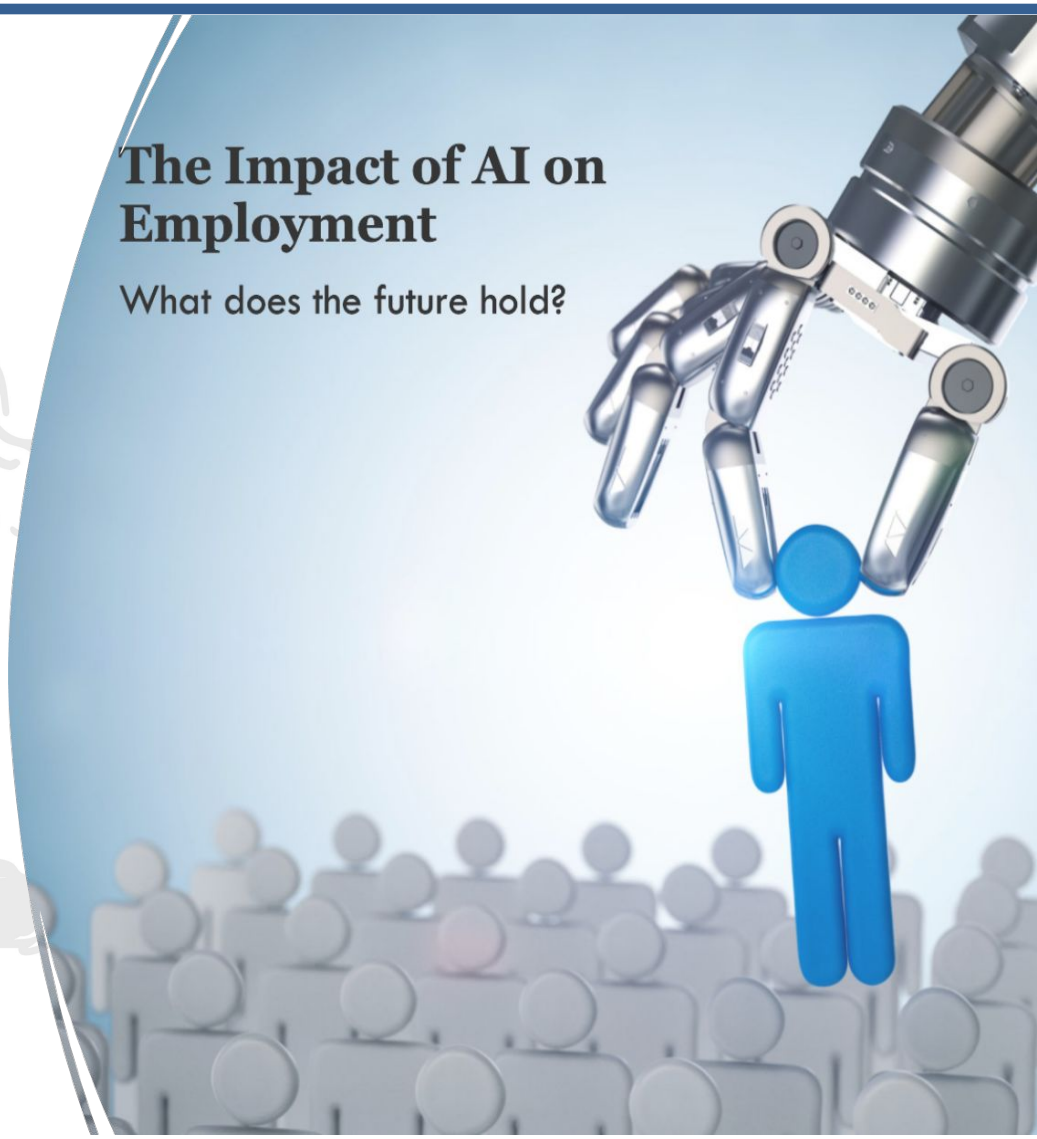
Deep Fakes and Misinformation

- **Manipulation of Media:** Spread misinformation, defame individuals, or manipulate public opinion.
- **Impact on Democracy:** Convincing fake news can undermine trust in media, influence elections.
- **Legal and Ethical Responses:** Developing legal frameworks and ethical guidelines.



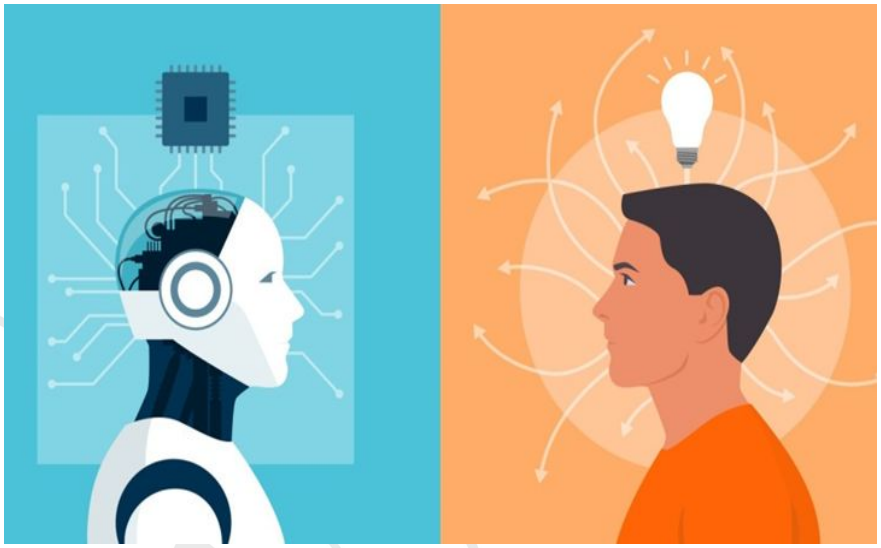
AI and Employment

- **Automation and Job Displacement:** Expected to displace many jobs
- **Reskilling and Education:** Ethical need to reskill and educate the workforce
- **The Gig Economy and AI:** The rise of AI-driven gig platforms.



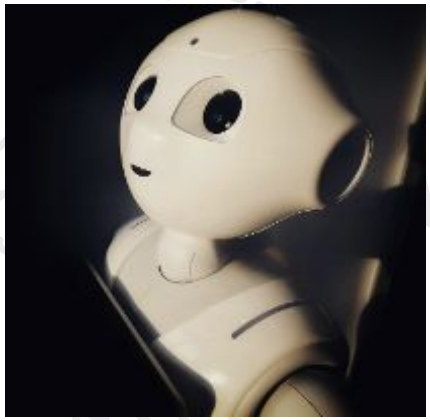
Human-AI Relationships

- **Dependency on AI:** Humans may become overly dependent
- **Emotional and Social Impact:** Potential for AI to affect mental health and social dynamics
- **AI and Human Identity:** Boundaries of identity will become increasingly important



FUTURE IMPLICATIONS AND PREDICTIONS

Increased Regulations



Ethical AI Development

Public Trust Importance

Balancing Innovation and Ethics

References

https://www.morganlewis.com/blogs/sourcingatmorganlewis/2024/01/ai-regulation-in-india-current-state-and-future-perspectives

https://www.unesco.org/ethics-ai/en?hub=32618

Lima G, Cha M, Jeon C, Park KS. The Conflict Between People's Urge to Punish AI and Legal Systems. Front Robot AI. 2021 Nov 8;8:756242. doi: 10.3389/frobt.2021.756242. PMID: 34820425; PMCID: PMC8606795.

https://www.geeksforgeeks.org/top-9-ethical-issues-in-artificial-intelligence/

https://www.coe.int/en/web/bioethics/common-ethical-challenges-in-ai