Responsible use of AI and Ethical Concerns

ONOJA ANTHONY, PH.D.

Introduction – The Big Shift

Artificial Intelligence (AI) is transforming industries, from healthcare to finance, education, and beyond, by enabling machines to perform tasks that typically require human intelligence. However, the rapid advancement and deployment of AI technologies raise significant concerns about their responsible use.

Objectives
✓ Understand the key risks and challenges associated with AI deployment.
✓ Identify ethical issues arising from AI applications.
✓ Explore proposed solutions to mitigate risks and promote ethical AI use.
✓ Foster critical thinking about balancing AI innovation with societal well-being.

Risks and Challenges of AI

AI systems, while powerful, pose several risks and challenges that can impact individuals, organisations, and society at large. These include technical, social, and systemic issues.

Bias and Discrimination

AI systems can inherit biases from training data, leading to discriminatory outcomes. For example, facial recognition systems have shown higher error rates for people with darker skin tones.

- ✓ **Impact:** Reinforces systemic inequalities, erodes trust in AI, and can violate legal and ethical standards.

- ✓ **Example:** In 2018, Amazon scrapped an AI hiring tool that downgraded resumes with female-associated terms because it was trained on male-dominated hiring data.

"Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer." – Cathy O'Neil

AI systems often rely on vast amounts of personal data, raising risks of unauthorised access, misuse, or surveillance.

Impact: Loss of individual privacy, potential for data breaches, and ethical concerns about consent.

The 2019 Cambridge Analytica scandal highlighted how AI-driven data analytics could manipulate voter behavior without transparent consent.

Lack of Transparency and Explainability

Many AI models, particularly deep learning systems, operate as "black boxes," making it difficult to understand their decision-making processes.

Impact: Hinders accountability, reduces trust, and complicates regulatory compliance.

Example: In healthcare, opaque AI diagnostics could lead to misdiagnoses without clear justification, risking patient safety.

Job Displacement and Economic Inequality

✓ Automation driven by AI can replace human labour in sectors like manufacturing, retail, and customer service.

✓ Automation driven by AI can replace human labour in sectors like manufacturing, retail, and customer service.

✓ Example: Studies estimate that AI could automate 20-30% of current jobs by 2030, particularly in routine-based roles.
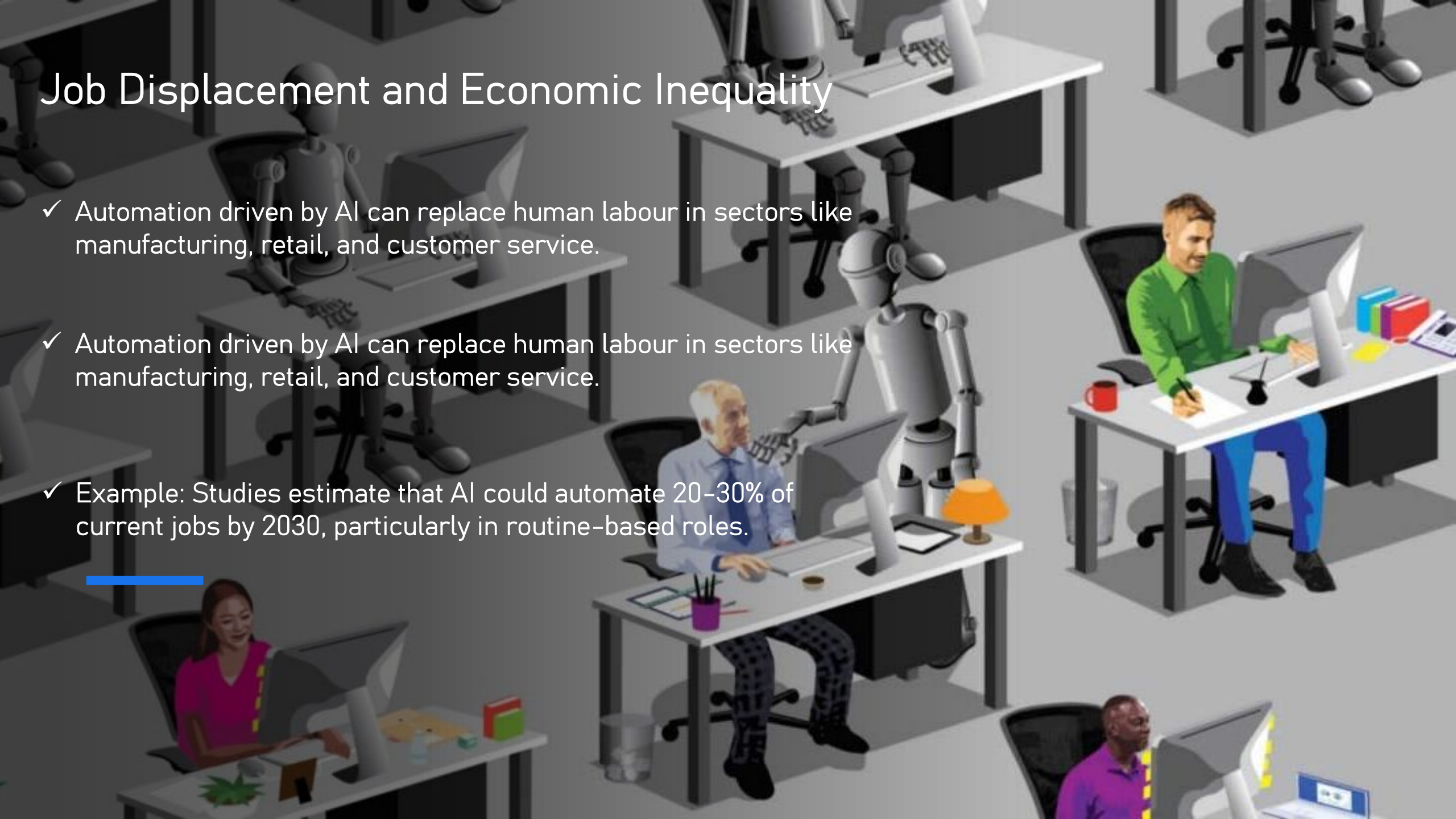
Security Risks

AI systems are vulnerable to adversarial attacks, where malicious inputs exploit model weaknesses, or to misuse in creating deepfakes or autonomous weapons.

- ✓ Impact: Threats to national security, public safety, and trust in digital content.

- ✓ Example: Deepfake videos have been used to spread misinformation, as seen in manipulated political speeches circulating online.

AI and Climate Change

Training large AI models requires significant computational resources, consuming vast amounts of energy.

Training large AI models (e.g., GPT-3, BERT) requires massive computational power, consuming thousands of kilowatt-hours (kWh).

Data centres powering AI models contribute significantly to greenhouse gas emissions, especially when powered by non-renewable sources.

Repeated training for hyperparameter tuning and large-scale inference increases this footprint.

Contributes to carbon emissions and environmental degradation.

Example: Training GPT-3 emitted an estimated 500 metric tons of CO_2—equivalent to a car driving over 1 million miles.

Model Size ≠ Intelligence: Larger models are not always better. There's often a diminishing return in accuracy vs. environmental cost.
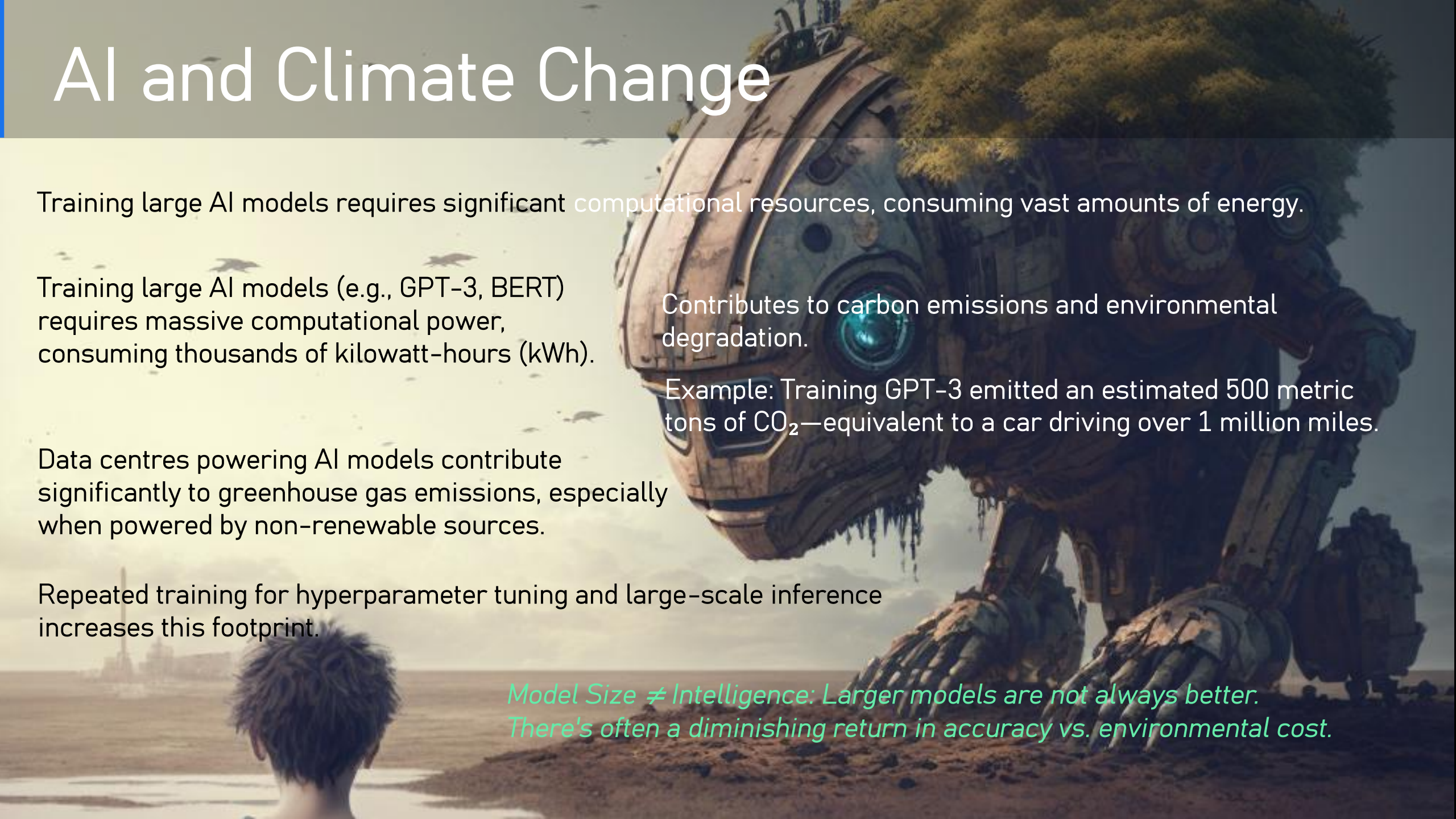
The risks and challenges of AI give rise to profound ethical questions that must be addressed to ensure fairness, accountability, and societal benefit.

Fairness and Equity: How can AI systems be designed to avoid perpetuating bias and ensure equitable outcomes for all demographic groups?

Accountability: Who is responsible when an AI system causes harm—developers, deployers, or users?

Autonomy and Consent: AI systems that make decisions impacting individuals (e.g., credit scoring, criminal justice) may undermine human autonomy.

Misuse and Dual-Use Concerns

Risks: Accidental escalation in conflict zones.
Lack of accountability for civilian harm or war crimes.
Arms race among nations to deploy faster, less accountable systems.

AI technologies are inherently dual-use, meaning they can serve both beneficial and harmful purposes depending on how they're deployed. This raises significant ethical and security concerns.

AI technologies like facial recognition, gait analysis, and predictive policing are increasingly deployed for mass surveillance, often without public knowledge or consent.

Autonomous Weapons (Lethal Autonomous Weapon Systems – LAWS)

AI can be embedded in drones, missiles, or robots capable of identifying and attacking targets without human intervention.

Examples: AI-guided drones with facial recognition or thermal imaging.
"Fire-and-forget" systems with no human-in-the-loop.

"Technologies developed for innocuous use (e.g., image classification) can be repurposed without your consent."

Transparency vs. Proprietary Interests

The tension between AI transparency and proprietary interests lies at the intersection of ethics, business strategy, and public accountability.

On one hand, transparency is essential for building trust, enabling accountability, and detecting potential harms like bias, unfairness, or safety risks. It involves practices like publishing model details, explaining decision-making processes, and sharing datasets or algorithms for peer review.

Many AI systems are developed by private companies that rely on intellectual property protection and trade secrets to maintain competitive advantage. Revealing the inner workings of models or datasets could compromise business models, enable copying, or expose them to reputational risk.

This leads to a conflict: the public and regulators increasingly demand openness about how AI affects users and society, while developers may resist sharing information that could threaten their commercial edge.

Case Studies

AI in Criminal Justice

Context: Predictive policing tools like COMPAS have been used to assess recidivism risk but faced criticism for racial bias.

Ethical Issue: Disproportionate impact on minority groups, undermining fairness.

Solution: Implement regular audits, transparent methodologies, and community input to ensure equitable outcomes.

Deepfakes and Misinformation

Context: Deepfake technology has been used to create misleading videos, eroding trust in media.

Ethical Issue: Threat to truth and public discourse.

Solution: Develop detection tools and promote media literacy to combat misinformation.

So What?

Proposed Solutions

To address the risks, challenges, and ethical issues of AI, a combination of technical, regulatory, and societal solutions is necessary.

Technical Solutions

✓ Bias Mitigation: Develop algorithms to detect and correct biases in training data, such as fairness-aware machine learning techniques.

Example: Use of tools like IBM's AI Fairness 360 to audit and mitigate bias in AI models.

✓ Explainable AI (XAI): Invest in research to make AI models more interpretable, enabling users to understand decision-making processes.

Example: Techniques like SHAP (SHapley Additive exPlanations) provide insights into model predictions.

✓ Robust Security: Implement defenses against adversarial attacks, such as input validation and model hardening.

Example: Google's Adversarial Training Toolkit enhances model resilience to malicious inputs.

Sustainable AI: Optimize algorithms to reduce energy consumption and use renewable energy for AI training.

Regulatory and Policy Solutions

✓ Global Standards and Regulations: Establish international guidelines for ethical AI development, such as the EU's AI Act, which classifies AI systems by risk level.

Example: The EU AI Act mandates transparency and human oversight for high-risk AI applications.

✓ Data Privacy Laws: Strengthen regulations like GDPR to ensure data is collected and used with explicit consent.

Example: GDPR's "right to explanation" empowers users to understand AI decisions affecting them.

✓ Accountability Frameworks: Create legal mechanisms to hold organisations liable for AI-related harms.

Example: Canada's proposed AIDA (Artificial Intelligence and Data Act) outlines penalties for non-compliant AI systems.

Societal and Organisational Solutions

✓ **Diverse Development Teams:** Promote diversity in AI research and development to reduce blind spots in design.
  Example: *Initiatives like Women in AI foster inclusive representation in the field.*

✓ **Public Awareness and Education:** Educate stakeholders about AI's capabilities and limitations to build informed trust.
  Example: *UNESCO's AI ethics training programs for policymakers and educators.*

✓ **Ethical AI Frameworks:** Adopt voluntary ethical guidelines, such as IEEE's Ethically Aligned Design, to guide AI development.

  Example: *Microsoft's Responsible AI Principles guide its AI product development.*

Collaborative Governance

✓ **Public-Private Partnerships:** Foster collaboration between governments, industry, and academia to align AI development with societal needs.

   Example: The Partnership on AI brings together stakeholders to advance responsible AI practices.

✓ **International Cooperation:** Harmonise AI ethics standards across borders to address global challenges like misinformation and autonomous weapons.

   Example: OECD's AI Principles promote trustworthy AI worldwide.

✓ *Supporting AI for good initiatives while resisting deployment in high-risk, rights-violating contexts.*

Hand-on
Session