Hello everyone,

We are Team C. Before we dive into our presentation, I would like to introduce my team members: Mr. Wing Lun Chan, Mr. Tsz Fung Sin, Mr. Waleed Mostafa, and I am Hamdan Ahmed.

Today, we will dive into the ethical challenges of AI and explore solutions to build trust in this transformative technology.

AI has brought remarkable benefits to industries. It enhances efficiency, saves time, and reduces costs, contributing significantly to the modern economy. However, these benefits come with profound ethical challenges. For example, as shown in this diagram, autogenerated comments improve efficiency but create accountability issues. This overlapping area highlights the ethical alignment we need to address for responsible AI development.

Challenges with AI can be categorized into three levels: short-term, medium-term, and long-term impacts.

Short-term challenges include privacy and transparency. Privacy concerns arise when users don't fully understand how their data is collected and used, leading to mistrust. Transparency issues occur when AI decisions are opaque, making it difficult for users to trust the systems.

In the medium term, we encounter challenges related to fairness, bias, and accountability. AI systems trained on biased data risk perpetuating inequalities based on race, status, or gender. Accountability issues arise when it's unclear who is responsible for AI decisions—developers, organizations, or the technology itself.

In the long term, risks include the erosion of human autonomy. Excessive reliance on AI could diminish individual decision-making, and the potential rise of AI superintelligence may surpass human control, leading to unpredictable consequences for society.

To better understand the importance of these challenges, let me share two case studies.

The first is the **Australian Values and Attitudes on Artificial Intelligence (AVA-AI)** study, which surveyed over 4,400 Australians on AI adoption in healthcare and social services. While 60% supported AI development, acceptance significantly decreased for applications with limited human involvement or lack of transparency.

This study measured ethical dimensions using a reverse-coded Likert scale, where lower scores represented greater importance. Results showed strong

consensus, with all dimensions scoring significantly below the midpoint of 3.0 on a 5-point scale.

Accuracy emerged as the most critical ethical dimension in both healthcare and social services, reflecting its fundamental importance. Fairness was highly valued, especially in social services, while human discretion was emphasized as essential in sensitive areas like healthcare. Speed, although moderately important in time-sensitive scenarios such as healthcare triage, ranked lower than accuracy and fairness. Reducing costs was considered the least important, indicating that public support for AI hinges more on ethical considerations than financial benefits. These findings highlight the need for AI systems to align with public expectations regarding ethical considerations.

The second case study examined **AI and Big Data across ten sectors**, including agriculture, government, insurance, and telecommunications. Using 22 stakeholder interviews, the study identified several universal concerns, with privacy and trust emerging as dominant issues, often driven by regulations like GDPR and public skepticism about data usage.

Sector-specific challenges included power imbalances in agriculture and justice concerns in manufacturing. Larger organizations, while having advanced AI implementations, faced greater scrutiny regarding ethical practices compared to smaller or public entities. Collaborations between public and private sectors often struggled to balance citizen-centric goals with profit motives.

Overall, these studies underscore the critical importance of addressing privacy, fairness, and transparency in AI deployment. They reveal a strong public demand for ethical AI systems, emphasizing the need for organizations to prioritize these dimensions in their practices.

Now that we've identified the challenges, let's discuss our proposed solutions.

We begin with existing standards. Organizations like IEEE, ISO, and OECD have introduced voluntary guidelines for ethical AI development. For instance, IEEE's P7000 standard addresses ethical concerns in system design, P7001 ensures transparency in autonomous systems, and P7002 focuses on robust data privacy practices. While these frameworks are not legally binding, they set a strong foundation for ethical AI. However, to address the full scope of challenges, we need more targeted and actionable measures.

Our approach tackles short-term, medium-term, and long-term challenges.

First, let's discuss short-term solutions, which address immediate concerns like privacy and transparency.

For Data Privacy, foundational steps include informed consent, where users clearly understand how their data is collected and used. Data minimization and purpose limitation ensure only the necessary data is collected and used for explicit purposes. Regular audits and privacy impact assessments help organizations identify and address risks before they escalate. Additionally, a clear data breach response plan builds user trust by ensuring swift action when issues arise.

Transparency is equally critical. Without transparency, how can users trust AI systems? To address this, we propose standardizing transparency requirements across AI applications. AI systems must also be explainable, meaning users should receive clear and jargon-free explanations for AI decisions. Comprehensive documentation of system design and user-centric feedback loops further enhance both trust and usability.

Moving to the **medium term**, accountability and bias must be addressed. For Accountability, it's crucial to define the scope and limitations of AI systems—what they can and cannot do. Additionally, we must identify all agents, both human and non-human, involved in the AI lifecycle. Clear responsibility for decisions and interrogation mechanisms for outputs are vital for ensuring humans remain in control as well.

Bias, another pressing issue, requires diverse data collection to reflect a wide range of demographics. Benchmarking procedures and regular audits help identify and mitigate biases, while re-weighting underrepresented groups ensures fair representation. These measures foster trust and prevent the perpetuation of inequalities.

Finally, in the **long term**, preserving human autonomy is essential. Enforcing regulation and ethical risk assessments should be conducted throughout the AI lifecycle to identify potential threats. Under continuous monitoring, AI systems should be designed to support and assist people in their tasks, rather than taking over or replacing them entirely. User-centric controls empower individuals to customize AI behaviour, and human override mechanisms guarantee that ultimate decision-making authority remains with people.

By addressing these concerns through short-term, medium-term, and long-term solutions, we can build AI systems that are trustworthy, equitable, and aligned with human values.

In conclusion, while AI offers immense potential to enhance industries, it also poses significant ethical challenges that we must address to build trust in this transformative technology. These challenges span across three critical levels:

short-term issues like privacy and transparency, medium-term concerns such as fairness, bias, and accountability, and long-term risks to human autonomy.

Our case studies highlight the public's demand for ethical AI systems. From Australians emphasizing accuracy and fairness over cost and speed to multi-sector insights revealing privacy and trust as universal concerns, the message is clear: ethical AI is not just a preference; it is a necessity.

To address these challenges, we proposed a structured, phased approach. In the short term, we focus on strengthening privacy and improving transparency. In the medium term, we tackle fairness and accountability by diversifying data and ensuring clear oversight. In the long term, we emphasize preserving human autonomy, with user-centric designs and mechanisms to keep humans in control.

By implementing these solutions, we can ensure that AI remains a tool that serves humanity, fostering trust, equity, and alignment with human values. Let us work collectively to make ethical AI the foundation of this technological revolution. Thank you!