Building Trust Between Humans and AI:
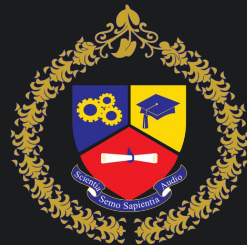
The Promise of Cooperative Artificial Intelligence

by

Professor Francisca Oladipo
Vice-Chancellor
Thomas Adewumi University
Nigeria

International Lecture Series 2025

Chandigarh Group of Colleges Jhanjeri, India

THOMAS ADEWUMI
UNIVERSITY,
OKO, KWARA STATE
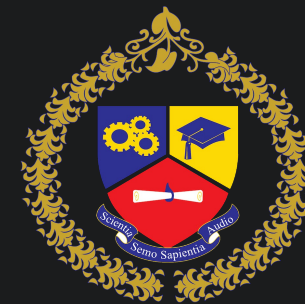Science | Technology | Medicine

CGC
JHANJERI

About Me

- **Professor of Computer Science**

  (Artificial Intelligence, Machine Learning, Natural Language Processing)

- **Administrator**

- **Vice-Chancellor, Thomas Adewumi University, Nigeria**

THOMAS ADEWUMI
UNIVERSITY,
OKO, KWARA STATE
Science | Technology | Medicine

About Thomas Adewumi University

Housekeeping

- Appreciation for the opportunity

- Gratitude for the cooperation between CGC and TAU

- Choice of a contemporary topic that will resonate with everyone

  - The matter of *trust* in the adoption of AI systems

  - Humans generally have trust issues, self-preservation, sceptics

- How can we achieve some level of trust in AI Systems?

The Current Landscape: AI's Rapid Advancement & the Trust Deficit

AI's Exponential Growth

From simple algorithms to complex machine learning models, AI is transforming every sector. This growth is fueled by increased data availability, computational power, and algorithmic innovation.

AI's capabilities now span from automating routine tasks to making critical decisions in healthcare, finance, and transportation.

The Trust Problem

Concerns about bias, job displacement, and lack of transparency are hindering AI adoption. Many people are wary of AI systems due to a lack of understanding of how they work and the potential for unintended consequences.

Addressing these concerns is crucial for fostering greater trust and acceptance of AI technologies.

The AI Trust Paradox:

Growing Adoption vs. Growing Concerns

We increasingly rely on AI while simultaneously harboring deep concerns

68% of organizations increased AI investments in 2024, while 71% of the public express concerns about AI risks

The "black box" problem undermines trust in high-stakes decisions For example, Deep learning models can contain billions of parameters with complex interdependencies,

"How can we trust what we cannot understand?"

Defining Trust in Human-AI Interactions

Trust as a multi-dimensional construct: <u>competence</u>, <u>reliability</u>, <u>fairness</u>, <u>transparency</u>, <u>alignment</u>
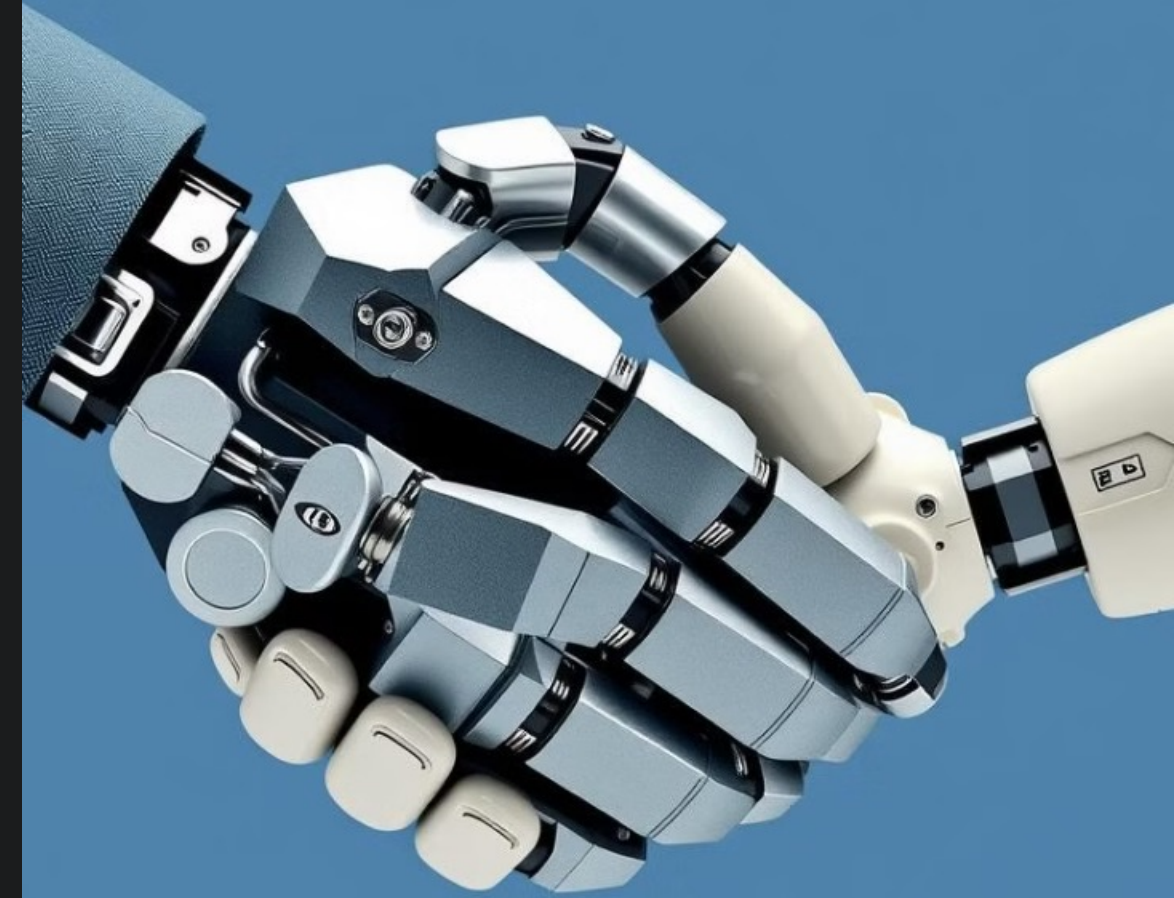
Competence: Can the AI System perform its intended function?

Reliability: Can the AI System perform consistently?

Fairness: Does it treat all users equitably?

Transparency: Can we understand how it works?

Alignment: Does it share our goals and values?

Defining Cooperative AI: A Human-Centered Approach

1   AI systems designed to work **with** humans, not replace them. This involves creating AI that complements human skills and abilities, enhancing productivity and creativity rather than rendering human roles obsolete. The focus is on synergy, where the combined capabilities of humans and AI exceed what either could achieve alone.

2   Emphasis on collaboration, mutual learning, and shared goals. Cooperative AI should facilitate seamless interaction between humans and machines, allowing them to learn from each other and adapt to changing circumstances. This requires AI systems to be transparent, explainable, and aligned with human values and objectives, ensuring that they work towards common goals in a trustworthy manner.

3   Contrasting with purely autonomous AI: Focus on augmentation, not full automation. Unlike autonomous AI, which aims for complete independence and decision-making authority, cooperative AI prioritizes human oversight and control. The objective is to empower humans with AI-driven insights and capabilities, enabling them to make better-informed decisions and take more effective actions, rather than relinquishing control to fully automated systems.

The Benefits of Cooperative AI: Enhanced Capabilities & Increased Trust

Combining AI's computational power with human intuition and creativity.

Human oversight mitigates risks associated with purely autonomous systems.

Trust in AI is higher when humans feel in control and understand how it works.

Case Studies: Cooperative AI in Action

AI-assisted diagnosis, personalized medicine, robotic surgery with human surgeons.

AI-powered tutoring systems that adapt to individual student needs, providing personalized feedback.

Collaborative robots (cobots) working alongside humans on assembly lines, improving efficiency and safety.

Challenges and Opportunities in Building Cooperative AI

Technical challenges   1

Developing AI systems that can effectively communicate and collaborate with humans.

2   Ethical and societal challenges

Addressing algorithmic bias and ensuring fairness.

Opportunities   3

Investing in research and development of Cooperative AI technologies.

Building Trust: Key Strategies for a Cooperative AI Future

1

Prioritizing transparency and explainability

Using XAI techniques to make AI decision-making processes more understandable.

2

Establishing accountability frameworks

Defining clear lines of responsibility for AI systems.
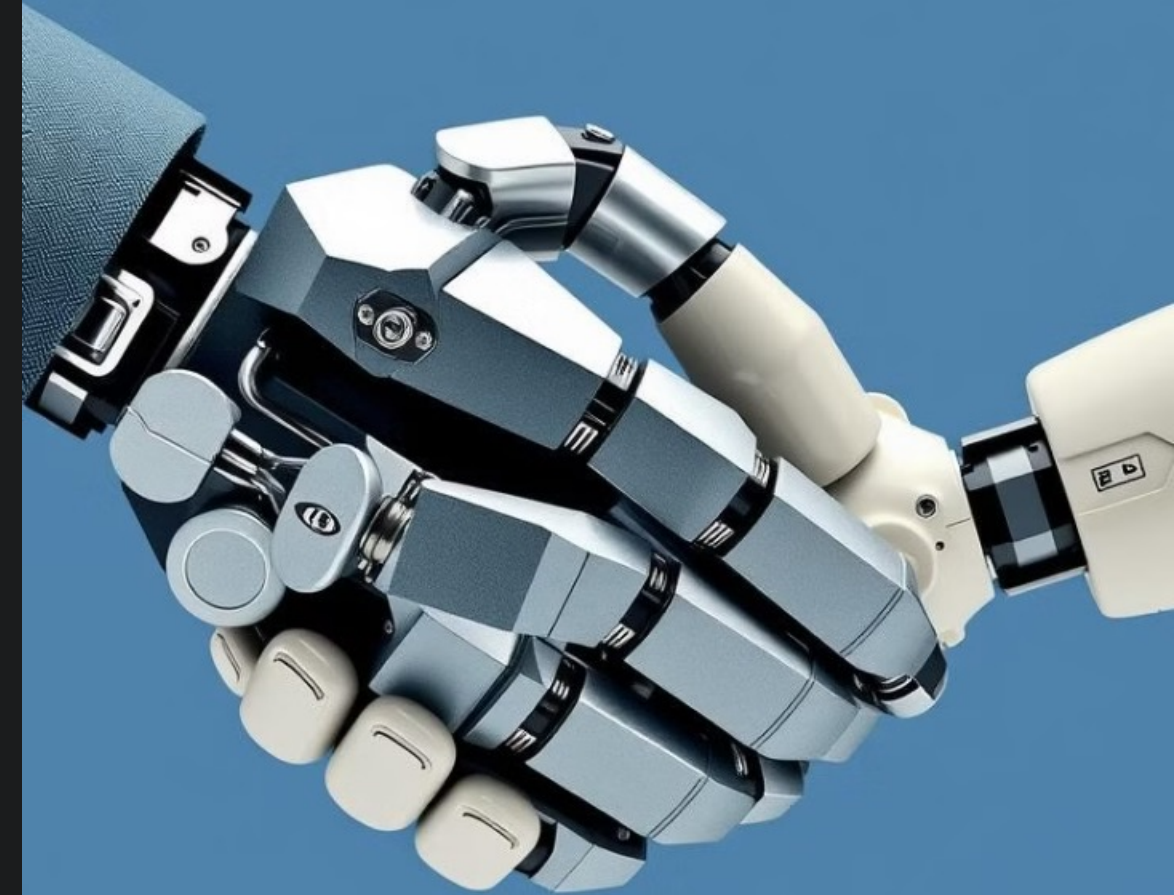
3

Promoting fairness and equity

Developing methods for detecting and mitigating algorithmic bias.
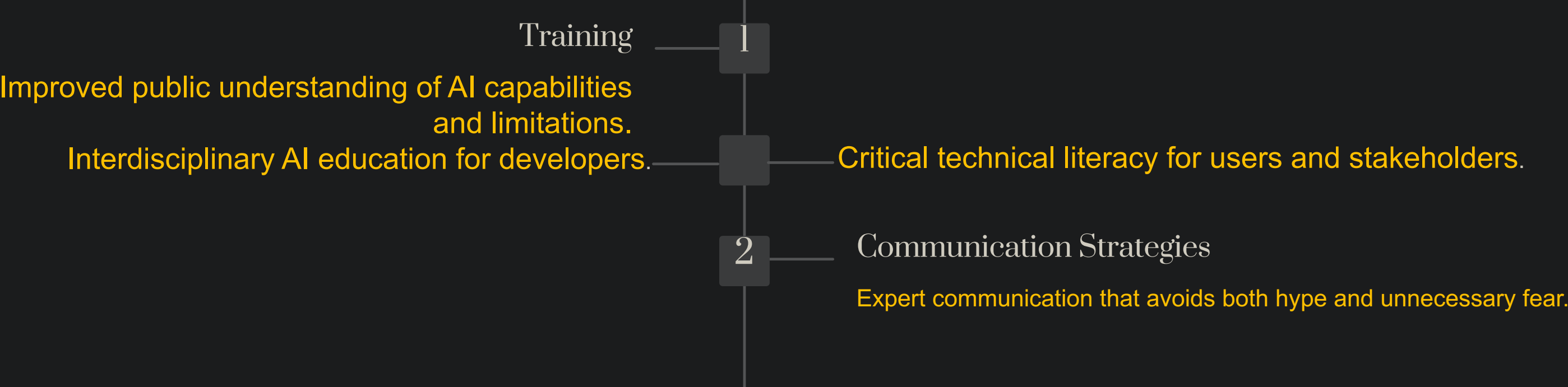
4

Engaging the public in dialogue

Hosting public forums and educational programs to discuss the implications of AI.

The Role of Educational Institutions

Training ——— 1

Improved public understanding of AI capabilities and limitations.

Interdisciplinary AI education for developers. ——— ▪ ——— Critical technical literacy for users and stakeholders.

2 ——— Communication Strategies

Expert communication that avoids both hype and unnecessary fear.

Conclusion

1 Trust is not just a "nice to have" but a functional necessity for effective human-AI collaboration

2 Technical solutions alone are insufficient; institutional and social dimensions are equally important
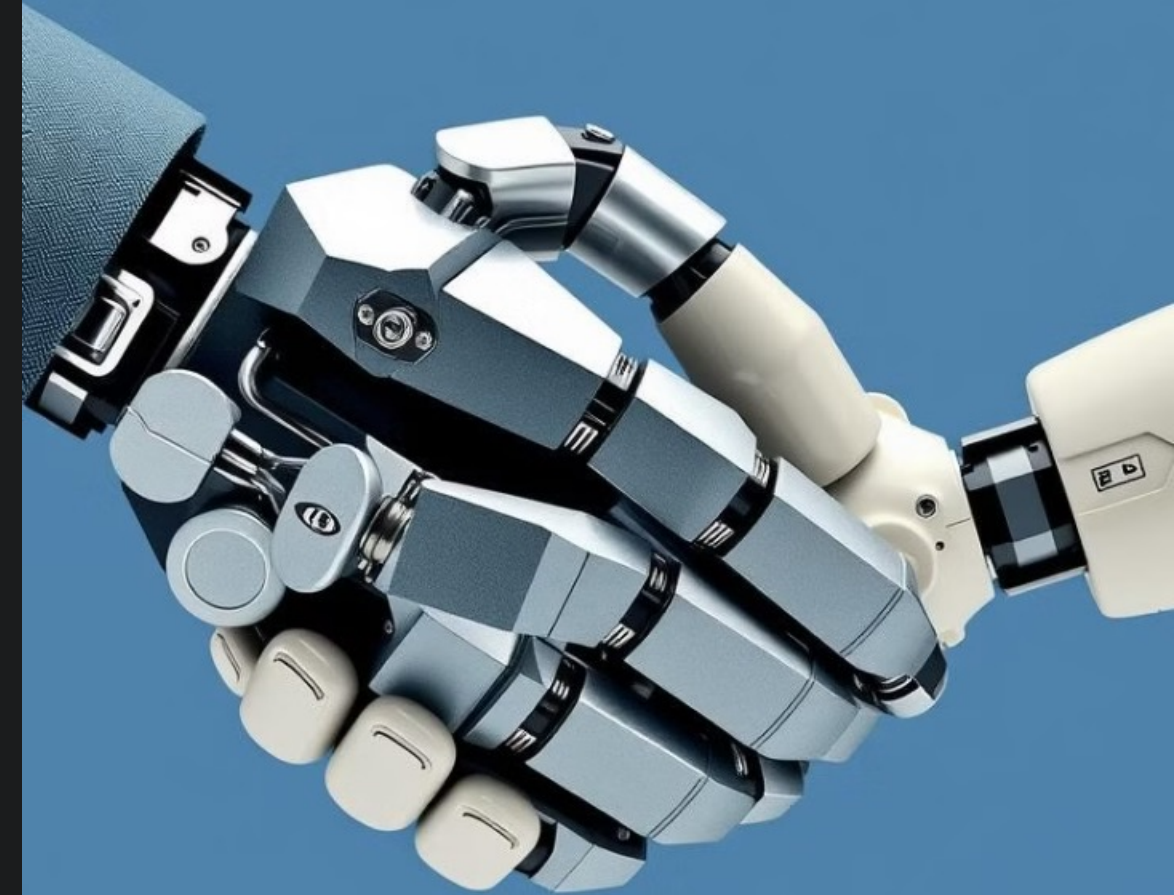
3 Technical solutions alone are insufficient; institutional and social dimensions are equally important
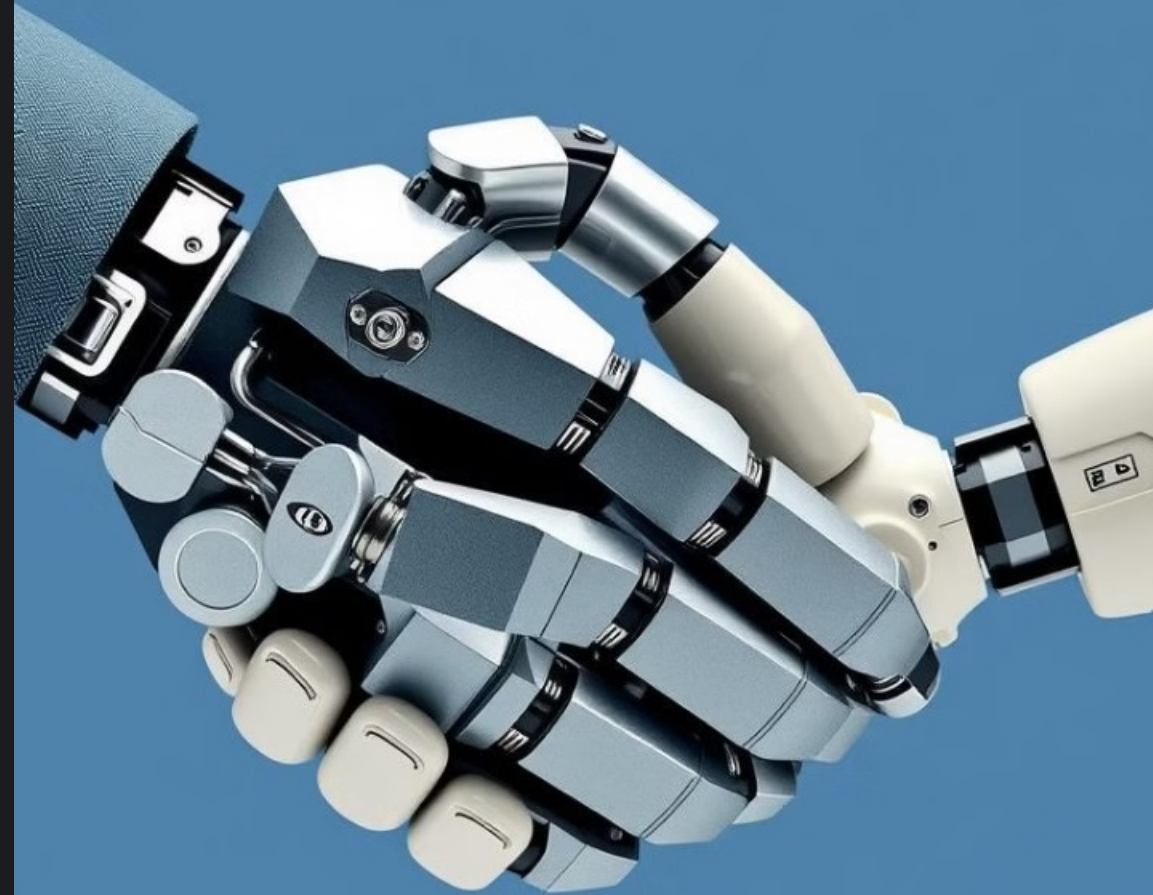
4 Cooperative AI offers a more comprehensive framework than intelligence-focused approaches

By prioritizing transparency, accountability, fairness, and public engagement, we can create a future where humans and AI work together to solve some of the world's most pressing challenges.

THANK YOU