Designing AI for Trust

AI: Intelligent Machines, Smart Policies
OECD Conference
27 October 2017

M-H. Carolyn Nguyen, Ph.D.
Director, Technology Policy
The **promise** of artificial intelligence

- **Education**
- **Healthcare**
- **Transportation**
- **Agriculture**
Human-Centred AI

AI Must:
1. Be designed to assist humanity
2. Be transparent
3. Maximise efficiencies without destroying the dignity of people
4. Be designed for intelligent privacy
5. Have algorithmic accountability so that humans can undo unintended harm
6. Guard against bias

Values People Must Bring to Bear:
7. Empathy
8. Education
9. Creativity
10. Judgment and accountability
We are pursuing AI to empower every person and every institution .. so that they can go on to solve the most pressing problems of our society and our economy.

– Satya Nadella
Ensure AI is Fair to All

Attract diverse pool of AI talent

Develop analytical techniques to detect bias

Develop guidelines for developing and deploying AI systems
Meaningful Transparency

Provide explanations of key elements of AI systems

Develop accurate and intelligible algorithms

Share best practices
Shaping development of trustworthy AI

Enable broad deployment & continued AI innovation with:

• Multi-stakeholder dialogues
• Practical principles for trustworthy AI
• Sharing of best practices
• Skills training
• Research into AI as part of solution
Thank you!