AI: Opportunities vs Risks
- Balancing Innovation and Regulations

Peter Pang , Managing Director of AA&T Consulting Services
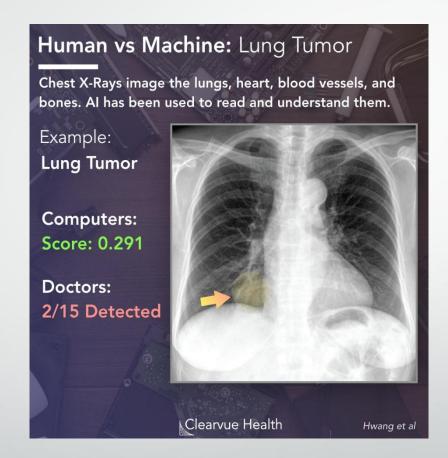
This or That:

Machine Radiologist or Human Radiologist

"AI and radiologist reports were not significantly different." (Oct 23)

(Source: US National Library of Medicine https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10556963/)

NHS rolled out new AI device with 99.8% effectiveness in detecting skin cancer in Aug 2024

(Source: https://www.england.nhs.uk/elective-care/best-practice-solutions/ai-based-skin-lesion-analysis-technology/ and https://www.medicaldevice-network.com/news/nhs-report-touts-effectiveness-of-skin-analytics-derm-ai-melanoma-tool/?cf-view)

Human vs Machine: Lung Tumor

Chest X-Rays image the lungs, heart, blood vessels, and bones. AI has been used to read and understand them.

Example:
Lung Tumor

Computers:
Score: 0.291

Doctors:
2/15 Detected

Clearvue Health   Hwang et al

AI: Opportunities vs Risks

This or That:

Robo taxi driver or Human taxi driver (2023)

(Source: Scientific America article:
https://www.scientificamerican.com/article/robo-taxi-takeover-hits-speed-bumps/

IEEE Spectrum article on AI risk of self driving cars:
https://spectrum.ieee.org/self-driving-cars-2662494269)
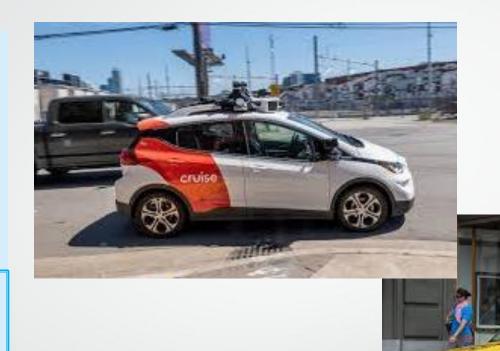
Update:

1. BBC: Driverless taxi* trying to win passengers' trust (Nov 24)
(https://www.bbc.com/future/article/20241115-how-robotaxis-are-trying-to-win-passengers-trust)

2. 1st fatal accident involving driverless car; Waymo not blamed (Jan 25)
(https://www.nbcbayarea.com/investigations/waymo-multi-car-wreck-san-francisco-driverless/3766860/)

3. Waymo vs Human drivers (29 Jul 25)
79% fewer airbag deployment; 93% fewer pedestrians' injuries
(https://waymo.com/safety/impact/)

* operated by Waymo (a google company) operating 150,000 rides per week in LA, SF and Pheonix.

AI: Opportunities vs Risks
- A brief history of AI

Physical Singularity

(Black hole)

vs

Technical Singularity
("Intelligence explosion" by an ultra-intelligent machine"; I.J. Good, 1965)

The Singularity Feedback Loop

Speculations Concerning the First Ultraintelligent Machine*
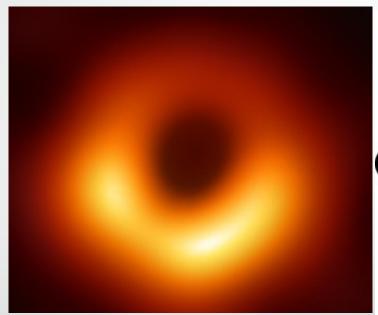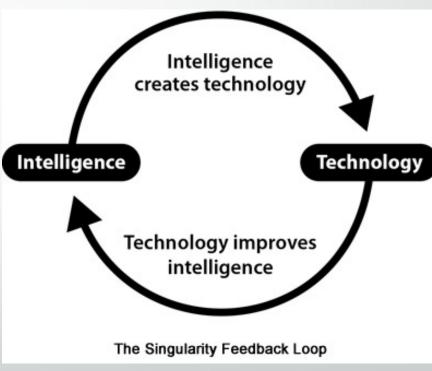
IRVING JOHN GOOD

Trinity College, Oxford, England and
Atlas Computer Laboratory, Chilton, Berkshire, England

1. Introduction 31
2. Ultraintelligent Machines and Their Value . . . 33
3. Communication as Regeneration 37
4. Some Representations of "Meaning" and Their Relevance to Intelligent Machines 40
5. Recall and Information Retrieval 43
6. Cell Assemblies and Subassemblies 54
7. An Assembly Theory of Meaning 74
8. The Economy of Meaning 77
9. Conclusions 78
10. Appendix: Informational and Causal Interactions . . 80
References 83

1. Introduction

The survival of man depends on the early construction of an ultra-intelligent machine.

"..an ultraintelligent machine could design even better machines; there would be an "intelligence explosion", and the intelligence of man would be left far behind." I.J. Good (1965)

AI: Opportunities vs Risks
- A brief history of AI

- *Moore's Law*

- *Real vs AI designed factory:*
  https://www.youtube.com/watch?v=xk1O2o6Fvbo

Real Factory vs Virtual Factory 2.0

AI: Opportunities vs Risks
- A brief history of AI

Launch of LLM models:

- *ChatGPT (OpenAI): Nov 22*
- *LLaMA (Meta AI): Feb 23*
- *Claude (Anthropic): Mar 23*
- *Grok (xAI): Nov 23*
- *Gemini (Google): Dec 23 (Bard, Google's former LLM Model: Mar 23)*
- *.......*

Launch of AI agent modules:

- *Gemini 2.0 for agentic era: Dec 24*
- *ChatGPT agent: Jul 25*
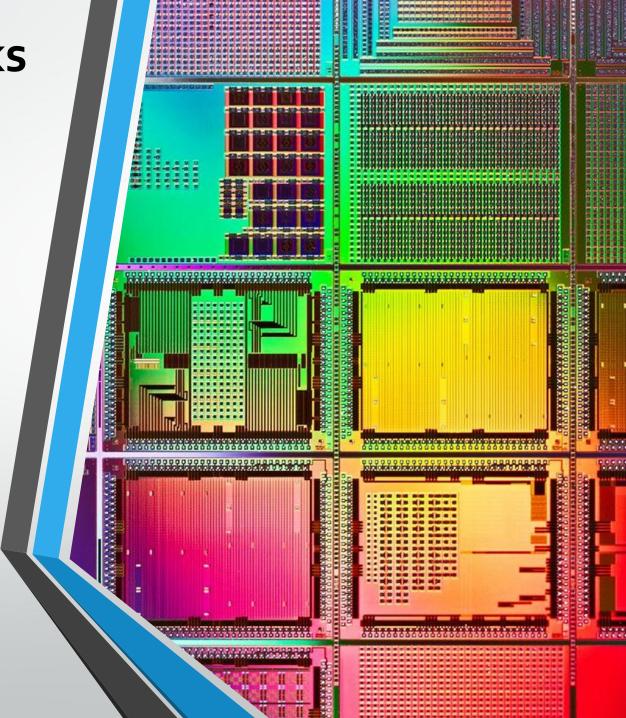- *.....*

AI: Opportunities vs Risks

- A brief history of AI

The 2024 Nobel Prizes: AI at the Forefront of Discovery

- The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks". https://www.nobelprize.org/prizes/physics/2024/popular-information/

- One half of the Nobel Prize in Chemistry 2024 went to Demis Hassabis and John Jumper of Google DeepMind who developed an AI model to solve a 50-year-old problem: predicting proteins' complex structures. https://www.nobelprize.org/prizes/chemistry/2024/press-release/

AI: Opportunities vs Risks
- A brief history of AI

Month and Year	Key Person/ Company	Key AI Development
May 1950	Alan Turing	**Turing Test** proposed as a method to determine if a machine can exhibit intelligent behaviours indistinguishable from a human.
1965	Irving John Good	**Concept of "Intelligence Explosion" (Technology Singularity):** Prediction that an ultraintelligent machine could design even better machines, leading to an exponential increase in intelligence.
April 1965	Gordon E. Moore (Intel co-founder)	**Moore's Law:** Observation that the number of transistors on an integrated circuit doubles approximately every two years, driving exponential growth in computing power.
1986	Geoffrey Hinton, David Rumelhart, Ronald Williams	**Backpropagation Algorithm:** Widespread popularization and practical application of the backpropagation algorithm for training neural networks, enabling multi-layer networks to learn complex patterns. *(Note: Geoffrey Hinton was awarded Nobel Prize in Physics 2024.)*
March 2016	DeepMind (Google)	**AlphaGo defeated Lee Sedol:** AlphaGo, an AI program, defeated the world champion Go player Lee Sedol 4-1, a significant milestone in AI's ability to master complex strategic games.
2024	Demis Hassabis and John Jumper of DeepMind (Google) and others	**Launch of AlphaFold:** In 2018 & 2020, AlphaFold beat other scientists in the protein structure prediction "Olympics", CASP. In 2021, Demis Hassabis, John Jumper and others published the AlphaFold 2 paper in Nature alongside open-source software and a searchable database. They won Nobel Prize in Chemistry 2024.
2022 -2023 /2024-2025	OpenAI, Google, Meta, Anthropic, xAI,	**Launch of LLM / Agentic models**

AI: Opportunities vs Risks
- A brief history of AI

Source of reference:

1. "Computing Machinery and Intelligence" by Alan Turing (Mind, 1950)
https://academic.oup.com/mind/article-abstract/LIX/236/433/986238

2. "Speculations Concerning the First Ultraintelligent Machine" by I.J. Good (Advances in Computers, 1965)
https://www.sciencedirect.com/science/article/abs/pii/S0065245808604180

3. "Cramming More Components onto Integrated Circuits" by Gordon E. Moore (Electronics Magazine, April 19, 1965)
https://www.cs.utexas.edu/~fussell/courses/cs352h/papers/moore.pdf
https://www.intel.com/content/www/us/en/history/virtual-vault/articles/moores-law.html

4. "Learning representations by back-propagating errors" by Rumelhart, Hinton, and Williams (Nature, 1986)
https://www.nature.com/articles/323533a0

4a. Neural Network Game: https://playground.tensorflow.org

4b. Nobel Prize in Physics 2024 Popular information page https://www.nobelprize.org/prizes/physics/2024/popular-information/

5. DeepMind Official Announcements: AlphaGo defeated Lee
https://deepmind.google/research/projects/alphago/

6a. "Highly accurate protein structure prediction with AlphaFold" by Demis Hassabis, John Jumper and others (Nature, 2021) https://www.nature.com/articles/s41586-021-03819-2

6b. "Accurate structure prediction of biomolecular interactions with AlphaFold 3" by Demis Hassibis and others (Nature, 2024) https://www.nature.com/articles/s41586-024-07487-w

6c. Nobel Prize in Chemistry 2024 Popular information page https://www.nobelprize.org/prizes/chemistry/2024/popular-information/

AI: Opportunities vs Risks

- AI Opportunities

Some AI transforming examples:

1. **Healthcare:**
   a. Disease detection and diagnosis
   b. Drug discovery and development
2. **Finance:**
   a. Fraud detection
   b. Risk assessment (Credit/Scenario analysis)
3. **Transportation/logistics:**
   a. Autonomous vehicles
   b. Route optimization
4. **Entertainment / Media:**
   a. AI-driven recommendation systems
   b. Smart/Creative content creation
5. **E-Commerce:**
   a. AI powered chatbot
   b. Customer Relationship Management tools
6. **Data security:**
   a. Threat detection and prevention
   b. Network monitoring
7. **Agriculture:**
   a. Crop / soil monitoring
   b. Agricultural robotics
8. **All industries**
   a. Administration automation
   b. Fraud prevention

AI: Opportunities vs Risks
- AI Opportunities

Some practical daily applications:

1. Project planning and role assignments
2. Draft text messages / emails / stories
3. Deep research
4. Calendar management
5. Audio/Podcast generation
6. Video generation

AI: Opportunities vs Risks

- AI Opportunities

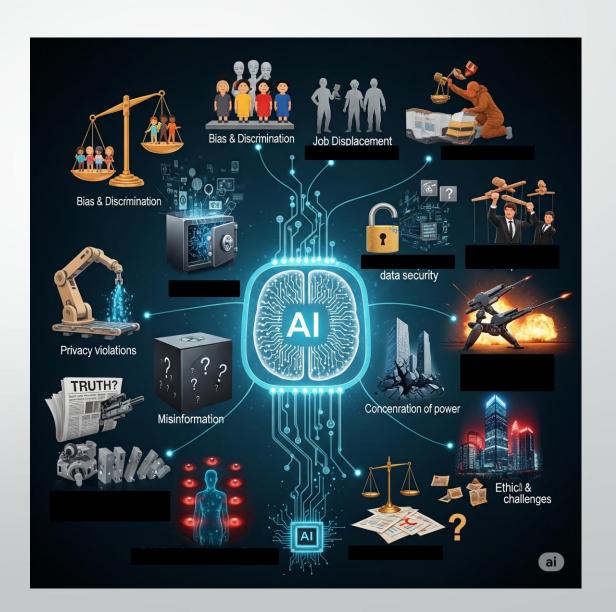**Cases of AI enabled productivity:**

1. Salesforce Einstein Service Agent
2. Amazon Alexa+
3. Bank of America "Erica"
4. Fraud Detection & Prevention: Mastercard, Visa, Revolut, DBS Bank

AI: Opportunities vs Risks
- AI Risks

Some Risks of AI development:

1. Bias and Discrimination

2. Job Displacement and Economic Inequality

3. Privacy Violations and Data Security

4. Lack of Transparency and Explainability (The "Black Box" Problem)

5. Misinformation and Manipulation

6. Autonomous Weapons and Escalation of Conflict

7. Concentration of Power

8. Over-reliance and Unintended Consequences

9. Existential Risks

10. Ethical and Regulatory Challenges

AI: Opportunities vs Risks
- Time to think

March 2023

"Godfather of AI" , Geoffrey Hinton

"We are moving into a period 1st time ever, we may have things more intelligent than us."

"There is enormous uncertainty about what's going to happen next; these things do understand; …. So, we need to understand what will happen next."

(An interview with 60 minutes, March 2023: https://www.youtube.com/watch?v=qrvK_KuleJk)

AI: Opportunities vs Risks
- Time to think

March 2023

Elon Musk is among the experts urging a 6 months halt to AI training in an open letter in March 2023 questioning the risk of loss of control of our civilization and suggesting a new and capable regulatory authorities dedicated to AI on oversight and tracking of highly capable AI systems.

(BBC 30 Mar 2023 https://www.bbc.com/ news/technology-65110030; Open letter: https://futureoflife.org/open-letter/pause-giant-ai-experiments/)

April 2023

Bill Gates says calls to pause AI won't 'solve challenges'

(Reuter 4 Apr 2023 https://www.reuters.com/technology/bill-gates-says-calls-pause-ai-wont-solve-challenges-2023-04-04/)

AI: Opportunities vs Risks
- Time to think

Mar 2018

Elon Musk: "Mark my words, AI is far more dangerous than nukes"

The billionaire tech entrepreneur said there needs to be a regulatory body overseeing the development of super intelligence.

(CNBC Mar 2018: https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html)

Jul 2025

" *…We are in **the intelligence Big Bang** right now. ..the most important thing is to be **maximize truth**…this Super Genius that ultimately will outsmart you, but you **can install the right values** and encourage it to be … truthful, honourable… Will it be good to humanity? I think it will…. even if AI ultimately proves bad for humanity, I'd at least like to be alive to see it happen."* Elon at the launch of Grok 4

(Fortune 10 Jul 2025 https://fortune.com/2025/07/10/elon-musk-xai-grok-tesla-optimus-mankind-humanity-robots/)

AI: Opportunities vs Risks
- Time to think

Sam Altman: (CEO of OpenAI)

Mar 2024

"I think you want a robust governance system over OpenAI or AGI"

(Lex Fridman Podcast #419 14 Mar 2024 https://www.youtube.com/watch?v=jvqFAi7vkBc)

May 2024

"Different types of AI systems will require different levels of regulation,"
(MIT Sloan 28 May 2024 https://mitsloan.mit.edu/ideas-made-to-matter/sam-altman-believes-ai-will-change-world-and-everything-else)

Jun 2024

"Global AI regulation shouldn't be overdone and cannot make it (human) safe."

(Bloomberg TV 23 Jun 2024 https://www.youtube.com/watch?v=-OmsJJ2RMZM)

AI: Opportunities vs Risks
- Time to think

Sam Altman: (CEO of OpenAI)

Apr 2025

*In a TED interview with Sam, he still stressed **the importance of a "preparedness framework" for safety** before releasing models. For the world he envisioned for his kid, he said, ". .... It'll be a world of incredible material abundance ... where the rate of change is incredibly fast and amazing new things are happening. And it'll be **a world where, like individual ability, impact, whatever, is just so far beyond what a person can do today.** I hope that my kids and all of your kids will look back at us (and say), 'They lived such horrible lives.'... "*

(**TED Interview with Sam Altman on 11 Apr 25 https://www.youtube.com/watch?v=5MWT_doo68k**)

Jul 2025

*Sam Altman warns of an **AI 'fraud crisis'** and said, "A thing that terrifies me is apparently there are still some financial institutions that will accept a voice print as authentication for you to move a lot of money ... (since) **AI has fully defeated most of the ways that people authenticate currently**, other than passwords."*

(CNN Business, 22 Jul 2025; https://edition.cnn.com/2025/07/22/tech/openai-sam-altman-fraud-crisis)

AI: Opportunities vs Risks

- Balancing Innovation vs Regulations

How to Balance between Innovation and Regulations ?

AI: Opportunities vs Risks
- AI Risks – Voluntary corporate rules
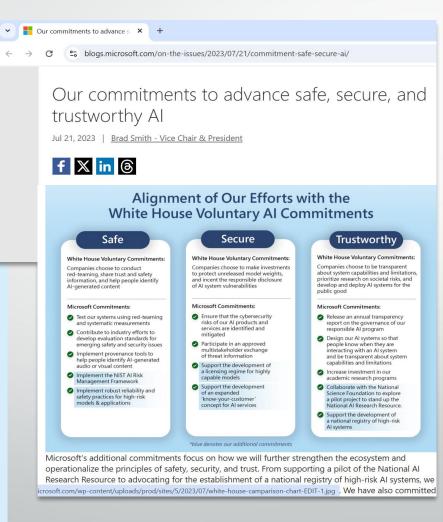
Corporate rules

Microsoft's commitment

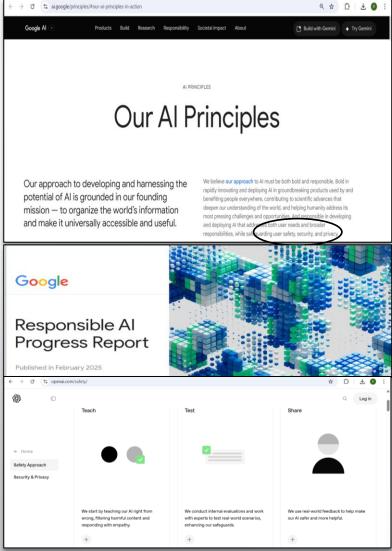https://blogs.microsoft.com/on-the-issues/2023/07/21/ commitment-safe-secure-ai/

Google's AI Principles

https://ai.google/responsibility/responsible-ai-practices/

OpenAI's safety and security policies

https://openai.com/safety/

AI: Opportunities vs Risks
- AI Risks – Regulatory regime

UNESCO's Recommendation on the Ethics of AI
https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

Law & regulations

EU AI Act: https://digital-strategy.ec.europa.eu/ en/policies/ regulatory-framework-ai

US: New Executive Order: "Removing Barriers to American Leadership in AI"

https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/
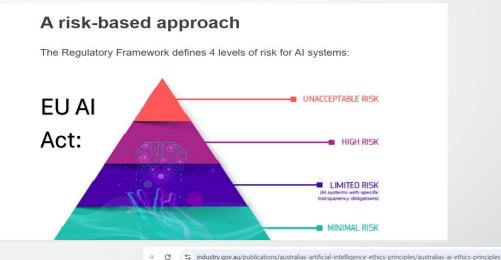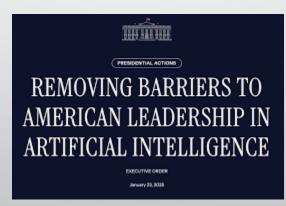
Voluntary rules:

Australia's AI Principles

https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles

A risk-based approach

The Regulatory Framework defines 4 levels of risk for AI systems:

EU AI Act:

UNACCEPTABLE RISK

HIGH RISK

LIMITED RISK (AI systems with specific transparency obligations)

MINIMAL RISK

Principles at a glance

- **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
- **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.
- **Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.
- **Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
- **Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.
- **Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.
- **Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.
- **Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

AI: Opportunities vs Risks
- More to consider

- Is AI just another Tech advancement? How can the (human) existential risk be minimised? What can we do to detect and monitor it?

- Innovation (AI or others): Is innovation always good for betterment of our life?

- Regulations: Are rules and regulations always necessary for new Tech? On what conditions, it should be imposed?

- Rules never eliminate risk. Detection, monitoring, education and training are always necessary.

Any more questions?

AA & T Consulting

Hope that you enjoy this session

If you need any help in technology or regulatory risk matters, please feel free to contact us by phone (+852 9181 8659 (HK); +61 452 371 753 (Aus.)), email (advisory@aathk.com) or via website's "contact us" page at: www.aathk.com or www.aataus.com .