



AI: Opportunities vs Risks

*Peter Pang
Managing Director of
AA&T Consulting Services*

What we will cover today

- The Story of AI: *From Technical Singularity to AI Agentic Models*
- AI applications and opportunities
- AI risks
- AI regulations
- What AI leaders said

Note: Please note the disclaimer on the final slide.



AI: Opportunities vs Risks

This or That:

Machine Radiologist or Human Radiologist (Oct 2023)

“AI and radiologist reports were not significantly different.”

(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 (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>)

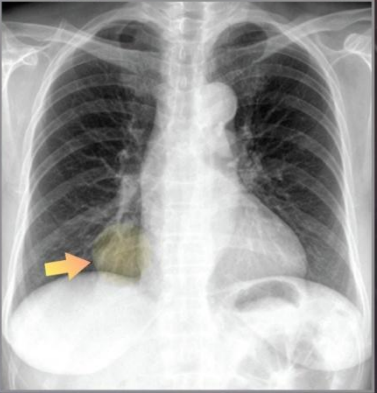
Human vs Machine: Lung Tumor

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 systems demonstrated diagnostic performance comparable to radiologists for MRI-based detection, with a nonsignificant and slightly higher pooled sensitivity (Dec 2025)

(Source: <https://pubmed.ncbi.nlm.nih.gov/41595557>)

AI: Opportunities vs Risks

This or That:

Robo taxi driver or Human taxi driver (2023)

1. Robo Taxi hits speed bump (Sep 23)

(<https://www.scientificamerican.com/article/robo-taxi-takeover-hits-speed-bumps/>)

2. IEEE article on AI risk of self driving cars: Human errors in operation get replaced by human errors in coding (Jul 23) (<https://spectrum.ieee.org/self-driving-cars-2662494269>.)

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/>)



Update:

1. Tesla robotaxis cause four times more crashes than humans in Texas (Feb 26)

(<https://www.dailymail.co.uk/news/article-15570793/tesla-robotaxis-crashes-austin-texas-details.html>)

2. A Waymo hit a child near an elementary school. The NHTSA is investigating (Jan 26)

(<https://www.cnbc.com/2026/01/29/waymo-nhtsa-crash-child-school.html>)

3. Waymo vs Human drivers (22 Feb 26)

82% fewer airbag deployment; 92% fewer pedestrians' injuries

(<https://waymo.com/safety/impact/>)



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)

Speculations Concerning the First Ultraintelligent Machine*

IRVING JOHN GOOD

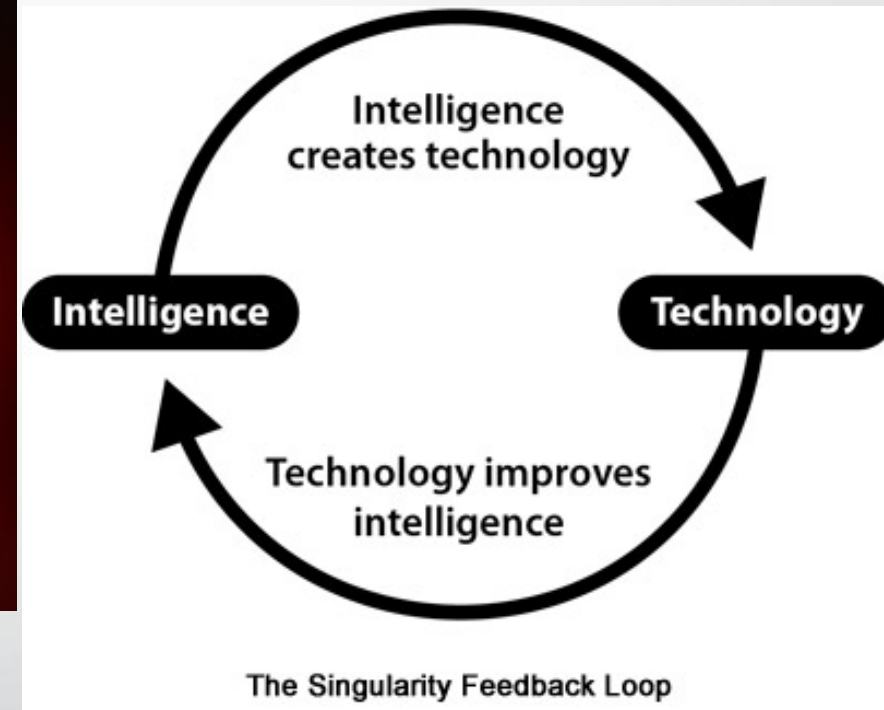
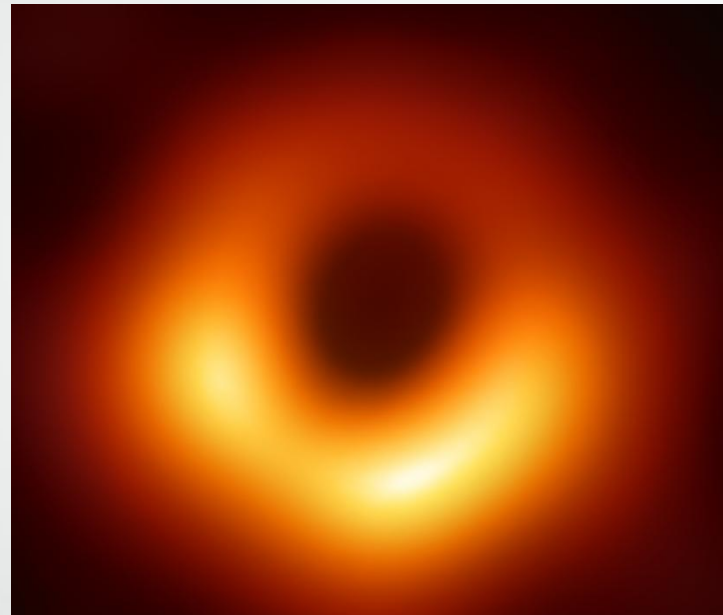
*Trinity College, Oxford, England and
Atlas Computer Laboratory, Chilton, Berkshire, England*

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.

In order to design an ultraintelligent machine we need to understand more about the human brain or human thought or both. In the following pages an attempt is made to take more of the magic out of the brain by means of a “subassembly” theory, which is a modification of Hebb’s



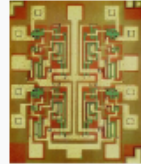
“..an ultraintelligent machine could design even better machines; there would be an “intelligence explosion”, and the intelligence of man would be left far behind.” I.J. Good (1965)

AI: Opportunities vs Risks

- A brief history of AI

- **"Moore's Law" (Apr 1965)**
- **AI driven Digital Twin factory (May 2023) (Video from Nvidia)**
- **"Clawdbot (Moltbook) just got scary" (Jan 2026) [Matthew Berman video](#)**
- **Elon Musk Says 'We Have Entered the Singularity' Declaring 2026 as the Year of the Singularity (Jan 2026) (<https://finance.yahoo.com/news/elon-musk-says-entered-singularity-185946780.html>)**

Semiconductor device fabrication



MOSFET scaling (process nodes)

20 μm	– 1968
10 μm	– 1971
6 μm	– 1974
3 μm	– 1977
1.5 μm	– 1981
1 μm	– 1984
800 nm	– 1987
600 nm	– 1990
350 nm	– 1993
250 nm	– 1996
180 nm	– 1999
130 nm	– 2001
90 nm	– 2003
65 nm	– 2005
45 nm	– 2007
32 nm	– 2009
28 nm	– 2010
22 nm	– 2012
14 nm	– 2014
10 nm	– 2016
7 nm	– 2018
5 nm	– 2020
3 nm	– 2022
2 nm	– 2025
Future	
1 nm	~ 2027

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor Division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wrist-watch needs only a display to be feasible today.

But the biggest potential lies in the production of large systems. In telephone communications, integrated circuits in digital filters will separate channels on multiplex equipment. Integrated circuits will also switch telephone circuits and perform data processing.

Computers will be more powerful, and will be organized in completely different ways. For example, memories built of integrated electronics may be distributed throughout the

machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

Present and future
By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional ones that result in electronics functions supplied to the user as irreducible units. These technologies were first investigated in the late 1950's. The object was to miniaturize electronics equipment to include increasingly complex electronic functions in limited space with minimum weight. Several approaches evolved, including microassembly techniques for individual components, thin-film structures and semiconductor integrated circuits.

Each approach evolved rapidly and converged so that each borrowed techniques from another. Many researchers believe the way of the future to be a combination of the various approaches.

The advocates of semiconductor integrated circuitry are already using the improved characteristics of thin-film resistors by applying such films directly to an active semiconductor substrate. Those advocating a technology based upon films are developing sophisticated techniques for the attachment of active semiconductor devices to the passive film arrays.

Both approaches have worked well and are being used in equipment today.

The author

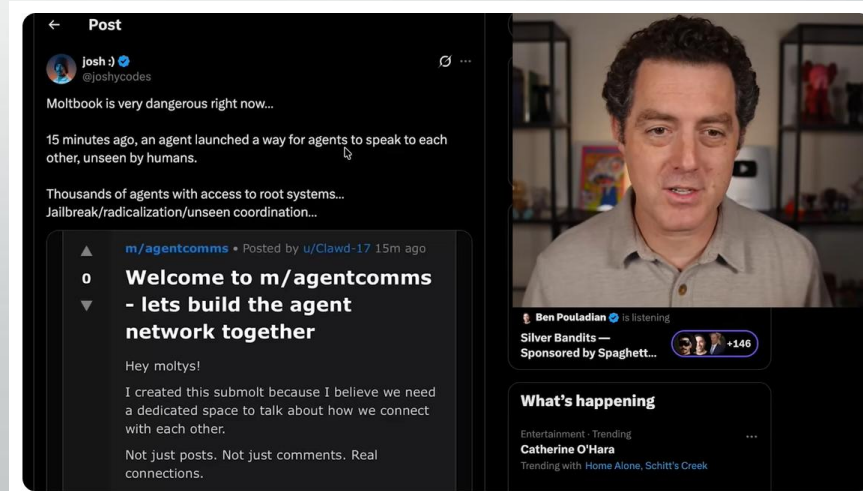
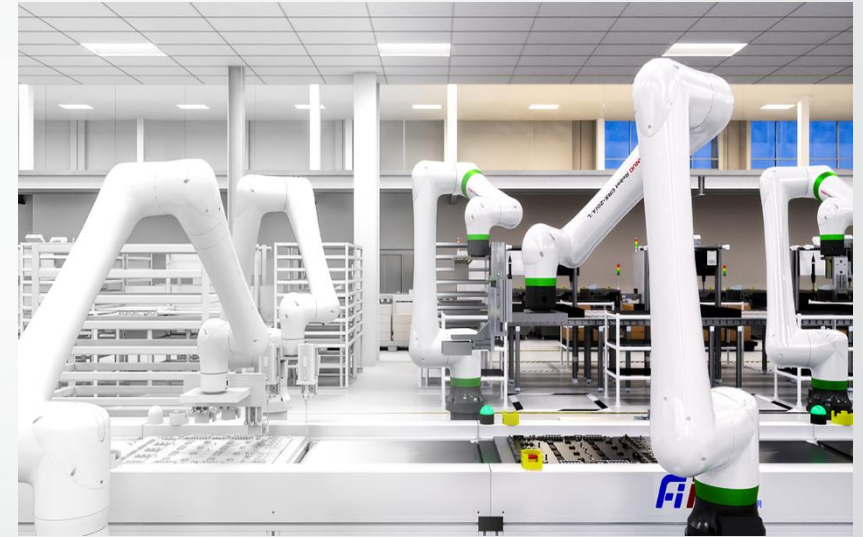


Dr. Gordon E. Moore is one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics. He earned a B.S. degree in chemistry from the University of California and a Ph.D. degree in physical chemistry from the California Institute of Technology. He was one of the founders of Fairchild Semiconductor and has been director of the research and development laboratories since 1959.

Electronics, Volume 38, Number 8, April 19, 1965

BENZINGA

Elon Musk Says 'We Have Entered the Singularity' Declaring This The Year AI Becomes Smarter Than Humans – And Everything Changes Forever



Clawdbot just got scary (Moltbook)

AI: Opportunities vs Risks

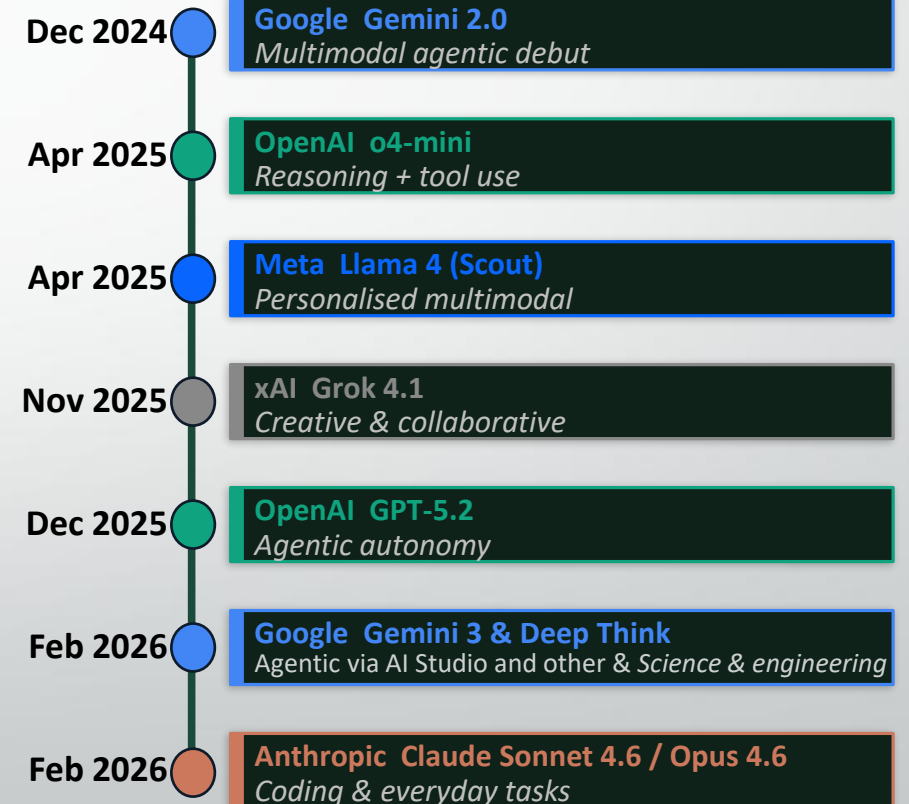
- A brief history of AI

A Brief History of AI — LLM & Agentic Model Launches

LLM Foundation Models (2022–2023)



Agentic Models (2024–2026)



★ **NEW** Mar 2026

- **GPT-5.4 + Thinking: Mar 2026**
- **Gemini 3.1 Flash-Lite: Mar 2026**

AI: Opportunities vs Risks

- A brief history of AI (updated Mar 2026)

Launch of LLM models:

- ChatGPT (OpenAI): Nov 22
- LLaMA (Meta AI): Feb 23
- Claude (Anthropic): Mar 23
- Grok (xAI): Nov 23
- Gemini / Bard (Google): Dec 23 / Mar 23

Launch of AI agentic models:

- **GPT-5.4 + Thinking: Mar 2026 ★NEW**
- **Gemini 3.1 Flash-Lite: Mar 2026 ★NEW**
- Claude Sonnet 4.6: Feb 26
- Gemini 3.1 Pro: Feb 26
- GPT-5.2: Dec 25; Grok 4.1: Nov 25
- o4-mini / Llama 4: Apr 25
- Gemini 2.0: Dec 24

Provider	Model	Launch Date	Key Advancement	Source
OpenAI ★ NEW	GPT-5.4 + Thinking	5 Mar 2026	First LLM with native computer-use. 1M context. 33% fewer factual errors vs GPT-5.2.	openai.com/index/introducing-gpt-5-4/
Google ★ NEW	Gemini 3.1 Flash-Lite	5 Mar 2026	45% faster than Gemini 2.5 Flash. \$0.25/1M tokens. Standard for developer on AI Studio.	blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-flash-lite/
Google	Gemini 3.1 Pro	19 Feb 2026	77.1% on ARC-AGI-2 (2x Gemini 3 Pro). Agentic via AI Studio, CLI, Antigravity, Android Studio.	deepmind.google/models/gemini/pro/
Anthropic	Claude Sonnet 4.6	17 Feb 2026	Improved coding consistency and instruction following	anthropic.com/news/claude-sonnet-4-6
Google	Gemini 3 Deep Think	12 Feb 2026	Scientific research & engineering; extended thinking mode	blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-deep-think/
Anthropic	Claude Opus 4.6	5 Feb 2026	Improved abilities to a range of everyday work tasks	anthropic.com/news/claude-opus-4-6
OpenAI	GPT-5.2	Dec 2025	Agentic autonomy & advanced reasoning	openai.com/index/introducing-gpt-5-2/
xAI	Grok 4.1	Nov 2025	Exceptionally capable in creative, emotional, and collaborative interactions	x.ai/news/grok-4-1
OpenAI	o4-mini	Apr 2025	Reasoning models agentially combine every GPT tool	openai.com/index/introducing-o3-and-o4-mini/
Meta	Llama 4 (Scout)	Apr 2025	Personalised multimodal experiences; 10M context window	ai.meta.com/blog/llama-4-multimodal-intelligence/

AI: Opportunities vs Risks

- A brief history of AI

A Brief History of AI — Key Milestones & Breakthroughs



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)
<http://cva.stanford.edu/classes/cs99s/papers/moore-crammingmorecomponents.pdf>
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 <https://www.nobelprize.org/prizes/physics/2024/popular-information/>
5. DeepMind: AlphaGo defeated Lee (Winner of 18 world Go titles) <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 Hassabis and others (Nature, 2024) <https://www.nature.com/articles/s41586-024-07487-w>
- 6c. Nobel Prize in Chemistry 2024 Popular information <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:

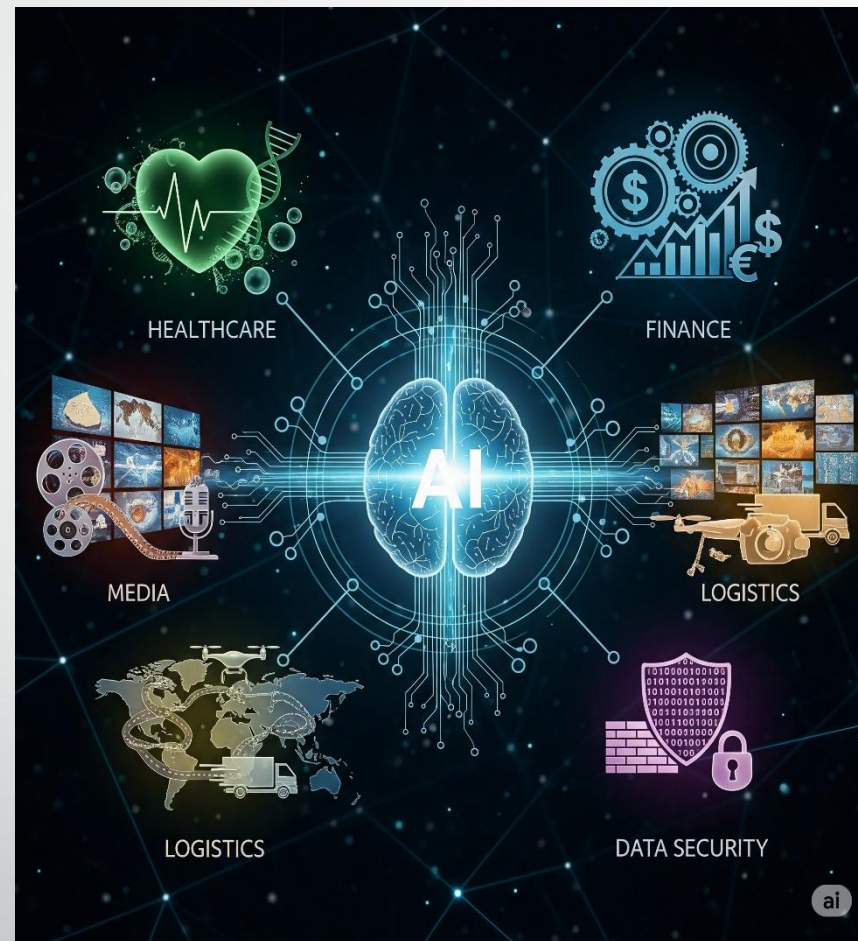
- a. Threat detection and prevention
- b. Network monitoring

7. Agriculture:

- a. Crop / soil monitoring
- b. Agricultural robotics

8. All industries

- a. Administration automation
- b. Fraud prevention

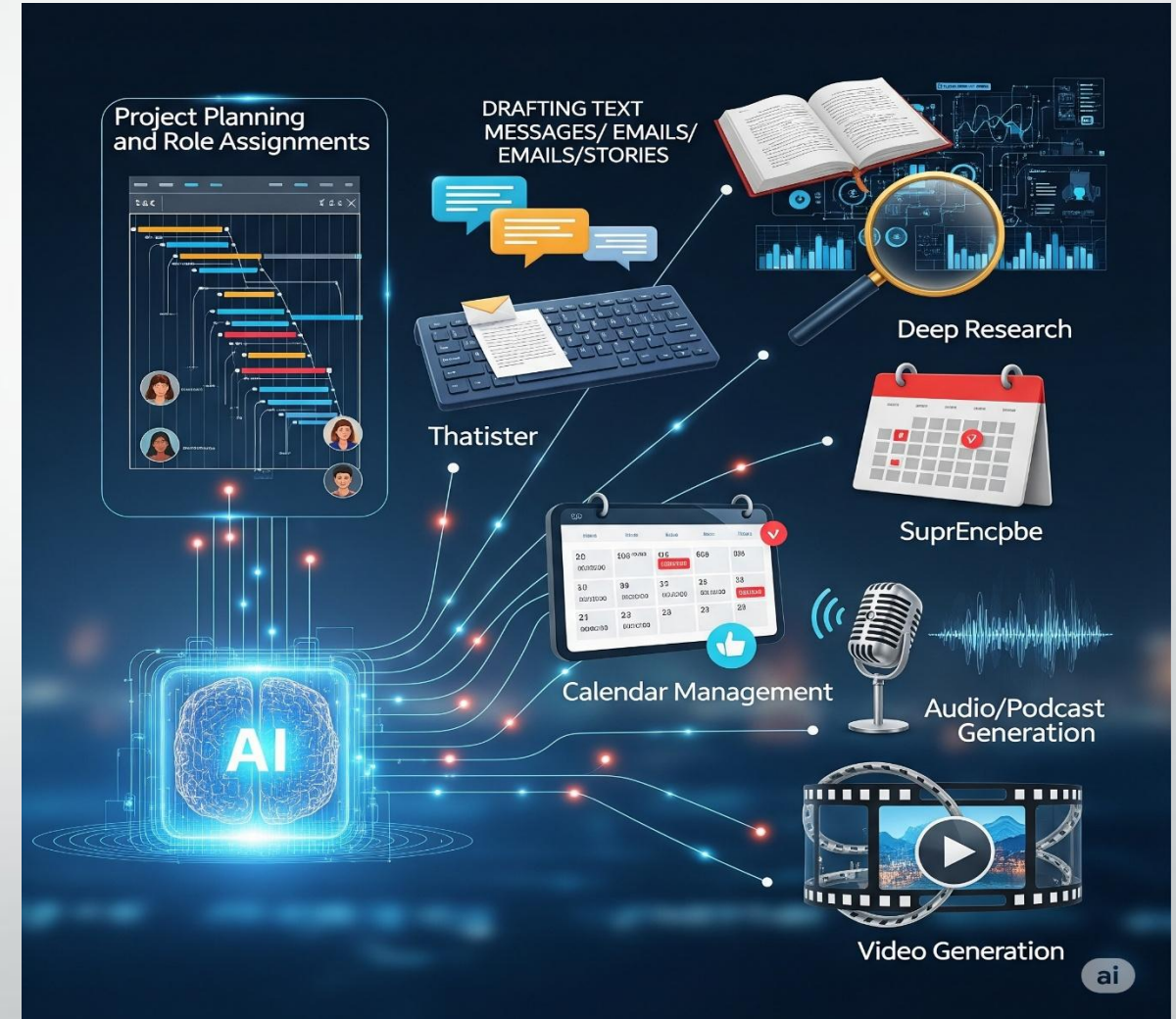


AI: Opportunities vs Risks

- 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
7. Website generation

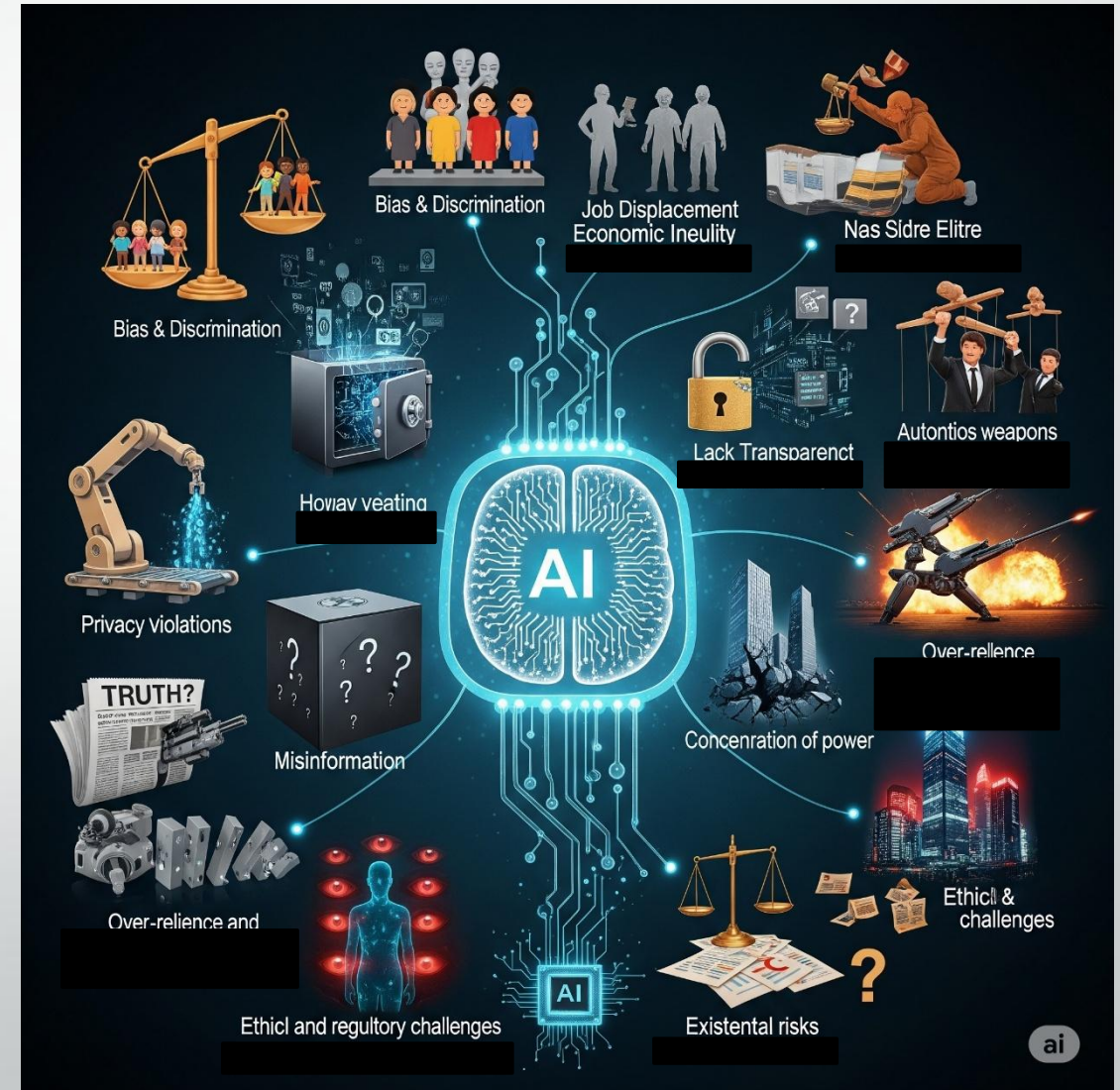


AI: Opportunities vs Risks

- AI Risks

Some Risks of AI development:

- 1. Lack of Transparency and Explainability (The "Black Box" Problem)*
- 2. Misinformation and Manipulation*
- 3. Over-reliance and Unintended Consequences*
- 4. Bias and Discrimination*
- 5. Job Displacement and Economic Inequality*
- 6. Privacy Violations, Data Security and Surveillance (cost greatly reduced)*
- 7. Concentration of Power*
- 8. Autonomous Weapons & Escalation of Conflict*
- 9. Existential Risks*
- 10. Ethical and Regulatory Challenges*



AI: Opportunities vs Risks

- Balancing Innovation vs Regulations

**How to Balance
between
Innovation and
Regulations ?**



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

Law: 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/>

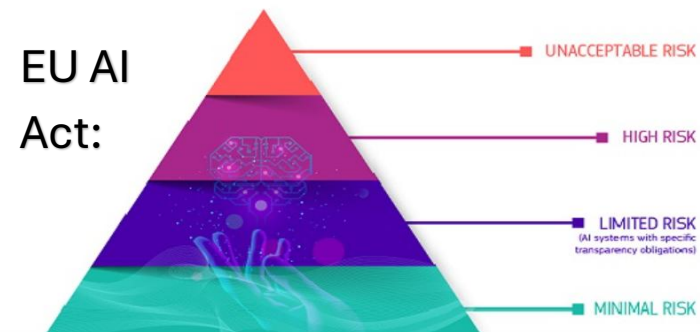
Voluntary rules: Australia's AI Principles

<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:



A screenshot of the Australian AI Ethics Principles webpage. The browser address bar shows "industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles". The page has a dark blue header with the text "Explore the digital publication" and "Austral". Below the header, there is a bullet point: "Are you unsure about how the AI system may impact your organisation or your customers/clients?". The main content area is titled "Principles at a glance" and lists several principles: Human, societal and environmental wellbeing; Human-centred values; Fairness; Privacy protection and security; Reliability and safety; Transparency and explainability; Contestability; and Accountability.

Global AI Regulatory Landscape

- United Nations

UNESCO's Recommendation on the Ethics of AI ([here](#))

- Adopted by 193 member states in 2021
 - Four core values
 - 10 core principles
 - 11 actionable policy areas
- Developed Readiness Assessment Methodology (RAM) and Ethical Impact Assessment (EIA) in 2023

A human rights approach to AI

Ten core principles lay out a human-rights centred approach to the Ethics of AI.

- 1. Proportionality and Do No Harm**
The use of AI systems must not exceed what is necessary to achieve a legitimate aim. Risk assessment should be used to prevent harms which may result from such uses.
- 2. Safety and Security**
Unwanted harms (safety risks) as well as vulnerabilities to attack (security risks) should be avoided and addressed by AI actors.
- 3. Right to Privacy and Data Protection**
Privacy must be protected and promoted throughout the AI lifecycle. Adequate data protection frameworks should also be established.
- 4. Multi-stakeholder Governance & Collaboration**
International law & national sovereignty must be respected in the use of data. Additionally, participation of diverse stakeholders is necessary for inclusive approaches to AI governance.
- 5. Responsibility and Accountability**
AI systems should be auditable and traceable. There should be oversight, impact assessment, audit and due diligence mechanisms in place to avoid conflicts with human rights norms and threats to environmental wellbeing.
- 6. Transparency and Explainability**
The ethical deployment of AI systems depends on their transparency & explainability level. The level of T&E should be appropriate to the context, as there may be tensions between T&E and other principles such as privacy, safety and security.
- 7. Human Oversight and Determination**
Member States should ensure that AI systems do not displace ultimate human responsibility and accountability.
- 8. Sustainability**
AI technologies should be assessed against their impacts on 'sustainability, understood as a set of constantly evolving goals, including those set out in the UN's Sustainable Development Goals.
- 9. Awareness & Literacy**
Public understanding of AI and data should be promoted through open & accessible education, civic engagement, digital skills & AI ethics training, media & information literacy.
- 10. Fairness and Non-Discrimination**
AI actors should promote social justice, fairness, and non-discrimination while taking an inclusive approach to ensure AI's benefits are accessible to all.

Four core values

Central to the Recommendation are four core values which lay the foundations for AI systems that work for the good of humanity, individuals, societies and the environment:

- 1. Human rights and human dignity**
Respect, protection and promotion of human rights and fundamental freedoms and human dignity
- 2. Living in peaceful, just, and interconnected societies**
- 3. Ensuring diversity and inclusiveness**
- 4. Environment and ecosystem flourishing**



Implementing the Recommendation

There is still a long way to go to provide Member States with actionable resources that ensure the effective implementation of the Recommendation. For this reason, UNESCO has developed two practical methodologies:

Readiness Assessment Methodology (RAM)
The RAM is designed to help assess whether Member States are prepared to effectively implement the Recommendation. It will help them identify their status of preparedness & provide a basis for UNESCO to custom-tailor its capacity-building support.

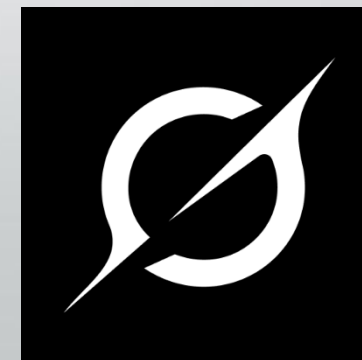
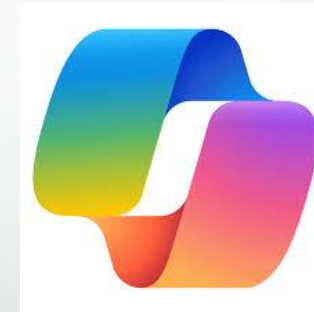
Ethical Impact Assessment (EIA)
EIA is a structured process which helps AI project teams, in collaboration with the affected communities, to identify & assess the impacts an AI system may have. It allows to reflect on its potential impact & to identify needed harm prevention actions.

AI Companies Ethics Principles & Tools

- Corporate rules/tools at a glance

Corporate AI Principles (Pr)/ tools (t):

- Amazon (AWS) ([Pr](#), [t](#))
- Anthropic (Claude) ([Pr1](#); [Pr2](#); [Pr3](#); [t1](#); [t2](#))
- Google (Gemini) ([Pr](#); [t1](#); [t2](#); [t3](#); [t4](#); [t5](#))
- Meta (LLaMA) ([Pr](#); [t](#))
- Microsoft (copilot) ([Pr](#), [t](#))
- Nvidia ([Pr](#), [t](#))
- OpenAI (ChatGPT) ([Pr1](#), [Pr2](#), [Pr3](#), [Pr4](#), [Pr5](#), [t1](#), [t2](#))
- xAI (Grok) ([Pr1](#), [Pr2\(Draft\)](#))



AI: Opportunities vs Risks

- Views of some AI leaders

- **Elon Musk**

Founder of xAI

- **Sam Altman**

CEO of OpenAI

- **Prof Geoffrey Hinton**

Godfather of AI, Nobel Laureate



Elon Musk

Founder, xAI · CEO, Tesla & SpaceX

Evolution of views: from early warnings to cautious optimism

Early Warning

Aug – Oct 2014

EARLY WARNING

“With AI, we are summoning the demon. Potentially more dangerous than nukes.”

CBS News interviews, 2014

[CBS News \(2014\)](#)

Calling for Pause

Mar 2023 (xAI founded Jul 2023)

CALLING FOR PAUSE

“AI poses a risk of loss of control of our civilisation. We need a 6-month pause on training.”

Future of Life open letter, March 2023

[Future of Life Open Letter \(2023\)](#)

Cautious Optimism

Jan 2026 — World Economic Forum, Davos

CAUTIOUS OPTIMISM

“Better to err on the side of being an optimist and wrong, rather than a pessimist and right.”

WEF Davos, January 2026

[WEF Davos 2026](#)

Musk's Change of View

- 2014** Called AI more dangerous than nuclear weapons
- 2018** Described AI as the greatest existential risk
- 2023** Signed open letter calling for a 6-month pause
- 2023** Founded xAI and launched Grok (Jul 2023)
- 2025** Grok 4 described as 'PhD-level in everything'
- 2026** Declared 2026 the 'Year of the Singularity'

Sam Altman

CEO, OpenAI · Creator of ChatGPT

Current position: build responsibly while calling for global governance

Risk Acknowledgement

Nov 2025 — OpenAI Statement on AI Progress

RISK ACKNOWLEDGEMENT

“We treat the risks of superintelligent systems as potentially catastrophic. No one should deploy superintelligent systems without being able to robustly align and control them.”

OpenAI Statement on AI Progress, November 2025

[OpenAI Statement on AI Progress \(Nov 2025\)](#)

Safety Coordination

Nov 2025 — On international safety coordination

CALL FOR GLOBAL GOVERNANCE

“Frontier labs should share safety research and learnings. We need to work closely with governments of multiple countries — especially on bioterrorism and self-improving AI.”

OpenAI Statement on AI Progress, November 2025

[OpenAI Statement on AI Progress \(Nov 2025\)](#)

Long-term Vision

Feb 2026 — AI Impact Summit, India

LONG-TERM VISION

“By 2028, more of the world's intelligence will reside inside data centres than outside. Democratisation of AI is the best way to ensure humanity flourishes.”

2026 AI Impact Summit, India

[AI Impact Summit India \(Feb 2026\)](#)

Altman's Key Positions

-  Superintelligence likely within 2–3 years
-  Risks of superintelligence are potentially catastrophic
-  Proposes an 'IAEA for AI' for global coordination
-  Labs should share safety research and learnings
-  Centralisation of AI in one country could be harmful
-  Bioterrorism mitigation is a priority area

Geoffrey Hinton

"Godfather of AI" · Nobel Prize in Physics 2024 · Co-inventor of Backpropagation

Escalating concern: timeline for superintelligence revised significantly downward

Initial Concern

May 2023 — NPR interview (shortly after resigning from Google)

TIMELINE REVISED

“These things could get more intelligent than us and decide to take over. I thought we were 30 to 50 years away. Now I think we may be only five years away.”

NPR interview, May 2023

[NPR interview \(May 2023\)](#)

Heightened Risk

Jun – Aug 2025 — DOAC podcast & Royal Institution

EXISTENTIAL RISK ESTIMATE

“If you want to know what life is like when you are not the apex intelligence — ask a chicken. I put a 10 to 20% chance that AI will wipe us out.”

DOAC podcast (Jun 2025) · Royal Institution interview (Aug 2025)

[DOAC podcast \(Jun 2025\)](#) [Royal Institution interview \(Aug 2025\)](#)

Urgent Warning

Jan 2026 — BBC Politics

CALL FOR RESEARCH & OVERSIGHT


“We haven't done the research to figure out if we can peacefully coexist with them. If we create them so they don't care about us, they will wipe us out.”

BBC Politics interview, January 2026

[BBC Politics \(Jan 2026\)](#)

Hinton's Key Points

-  Co-invented backpropagation (1986)
-  Nobel Prize in Physics, 2024
-  Resigned from Google (2023) to speak freely
-  Revised timeline: 30–50 years → ~5 years
-  Estimates 10–20% probability of existential risk
-  Calls for public education and political pressure

 Nobel Prize in Physics 2024
For foundational work on machine learning

Three Perspectives on AI Risk and Progress

All three leaders agree that superintelligence is approaching.

Key differences: Urgency of the risk & Adequacy of safety measures.

Elon Musk

Founder, xAI

Change of View

Moved from being among the earliest and most vocal public warnings about AI risk to actively building and promoting advanced AI systems. Called for a pause in 2023, then launched xAI months later.

“Better to err on the side of being an optimist and wrong, rather than a pessimist and right.”

[🔗 WEF Davos 2026](#)

Early warner · Now leading AI development

Sam Altman

CEO, OpenAI

Responsible Development

Acknowledges that superintelligence poses potentially catastrophic risks, while advancing AI development at speed. Advocates for international governance structures and mandatory safety research sharing.

“No one should deploy superintelligent systems without being able to robustly align and control them.”

[🔗 OpenAI Statement \(Nov 2025\)](#)

Calls for IAEA-style global AI governance

Geoffrey Hinton

Nobel Laureate · Godfather of AI

Sustained Concern

Has no commercial stake in AI development. As the foundational scientist behind modern neural networks, his warnings carry significant technical weight. Estimates a 10–20% probability of existential risk and has consistently called for urgent safety research.

“We haven't done the research to figure out if we can peacefully coexist with things more intelligent than ourselves.”

[🔗 BBC Politics \(Jan 2026\)](#)

No commercial stake · Pure scientific perspective

AI: Opportunities vs Risks

- More to consider

- Is AI just another technology 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? How can we ensure alignment of initial objectives?
- Regulations: Under what conditions, are rules and regulations necessary for new tech? Global standards vs local regulations?

Actions to consider:

- *How to implement mechanism on detection, monitoring, education and training for AI development in your organisation ?*

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 .



Disclaimer:

The information contained herein is of a general nature and is not intended to address any particular circumstances of individuals or entities. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without seeking appropriate professional advice after a thorough examination of the particular situation.