<u>**Generative AI Hackathon for SDGs**</u>

Designing your AI Solution Ethically

Brian W Tang

Founding Executive Director, LITE Lab@HKU

bwtang@hku.hk

October 4, 2024

The University of Hong Kong
LITELAB
Law | Innovation | Technology | Entrepreneurship

Generative AI Hackathon for SDGs

GenAI Hackathon for SDGs

21 Sep - 19 Oct 2024

Judging Criteria

Teams must demonstrate the following in their pitches:

- Proper scoping of social problem (what is the pain point?)
- Convincing and sustainable business model
- Sufficient application of GenAI technology
- Team competency to implement solution

Feasibility and impact   **Novelty and innovation**   **Transdisciplinary co-creation**   **Sustainability and ethics**

- Business plan that shows scalability, adaptability, and potential integration with existing systems or processes

- Potential replication of the solution in different contexts

- Awareness and consideration of potential ethical implications and risks such as bias, privacy concerns, or unintended consequences.

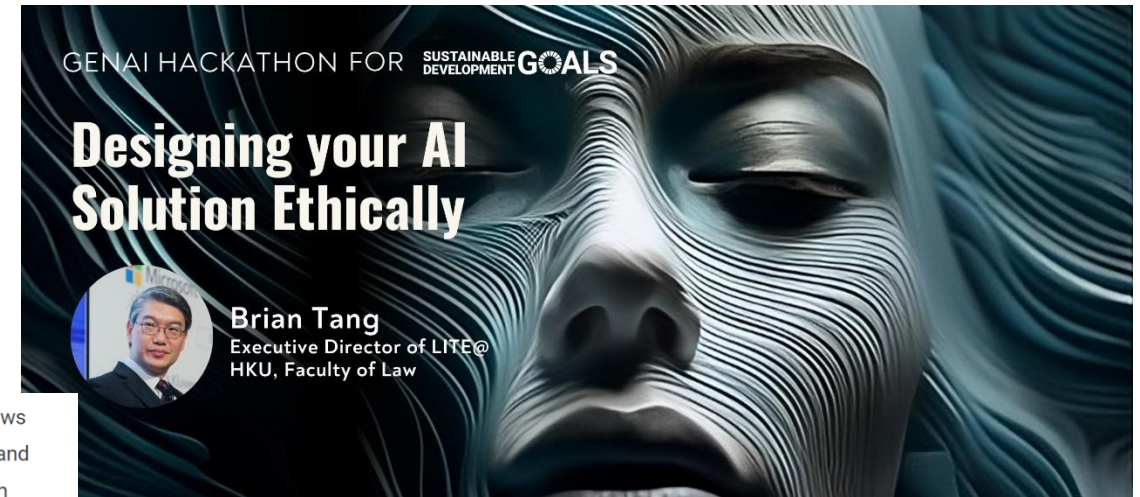GENAI HACKATHON FOR SUSTAINABLE DEVELOPMENT GOALS

Designing your AI Solution Ethically

Brian Tang
Executive Director of LITE@ HKU, Faculty of Law

Hack4Good 2024: Designing your AI Solution Ethically

As more technology startups and social entrepreneur founders use GenAI to create impactful solutions, given the rising appreciation of the externalities and new regulations globally, it has become increasingly important for AI ethics and governance be an integral part of the design process. This 1 ½ hour workshop introduces Hack4SDG participants to salient AI ethical design considerations when creating and implementing their SDG solutions to address our aging population, sustainable food and waste systems and inclusive cities for this hackathon, and beyond.

Date: October 4, 2024 (Fri)
Time: 4:30 pm – 6:00 pm
Venue: [CHANGED TO ZOOM ONLY]

Brian W Tang - Legal Innovation, Technology and Impact Journey

LITE LAB@HKU IS AN INTERDISCIPLINARY PROGRAM BY HKU LAW FACULTY

Chapter Co-author, *Fintech Book* (2017), *Regtech Book* (2019), *Legaltech Book* (2020), *AI Book* (2020) Artificial Intelligence In Finance (2023)

Organised HK's 1st Legaltech & Regtech Hackathon (2018)

Lead instructor of world's largest Fintech MOOC(2018) >177,000 students

Law, Innovation, Technology & Entrepreneurship Lab (LITE Lab) at University of Hong Kong, Faculty of Law - Founding executive director (2019 - present); HKU-SCF Fintech Academy

ACMI – Managing director (2014 – present)

Credit Suisse - Investment Banking Division Executive Director (2004 – 2014) China IBD JV; co-chair of Hong Kong Charity Committee; Impact Investing (CIC - PB); Microfinance Advocates

Sullivan & Cromwell (NY, CA) - Senior Associate (1997– 2004)

Mallesons (Perth) - Solicitor (1993 – 96) [+ NYU LLM 1996-7]

Volunteer Association Leadership Roles
Global Alliance of Impact Lawyers – APAC Board (2022–present)
Fintech Association of Hong Kong - Board (2021–present); Founding Regtech Committee Co-Chair (2018-21)
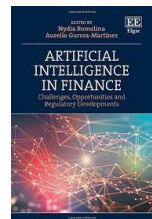
Asia Pacific Legal Innovation & Technology Association – Co-chairperson (2021 – present); Steering Committee (2019 -21)

Association of Corporate Counsel (Hong Kong) – Executive Committee (2009 – 14)

Designing your AI Solution Ethically

Generative AI Hackathon for SDGs

University
7 responses

- HKU
- City U
- HKBU
- HKUST

14.3%
28.6%
14.3%
42.9%
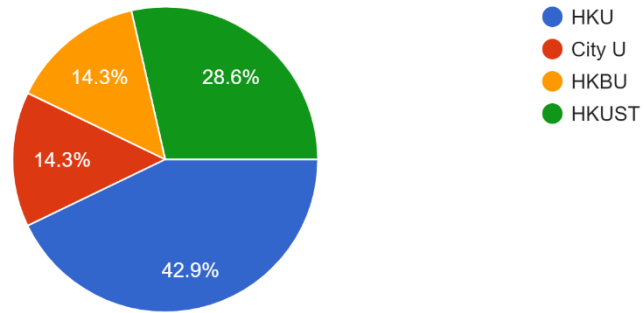
Thematic Tracks

Good Life for an Ageing Population

Innovating for enhanced, fulfilling senior lifestyles.

Sustainable Food and Waste Systems

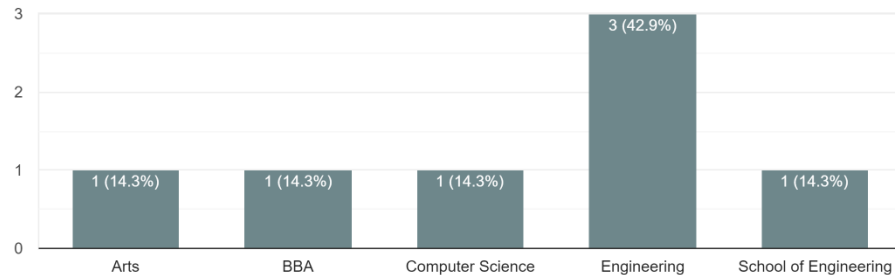Revolutionising practices for eco-friendly consumption and waste reduction.

Inclusive City for All

Reimaging the city that welcomes and empowers everyone.

SDG Thematic Track
7 responses

- Good Life for an Ageing Population
- Sustainable Food and Waste Systems
- Inclusive City for All

71.4%
14.3%
14.3%

Faculty
7 responses

1 (14.3%)	1 (14.3%)	1 (14.3%)	3 (42.9%)	1 (14.3%)
Arts	BBA	Computer Science	Engineering	School of Engineering
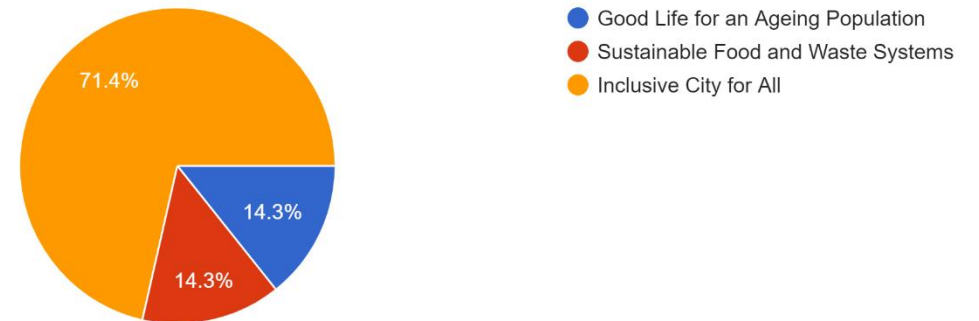
Designing your AI Solution Ethically

Understanding Hong Kong: "One Country Two Systems" and "Greater Bay Area"
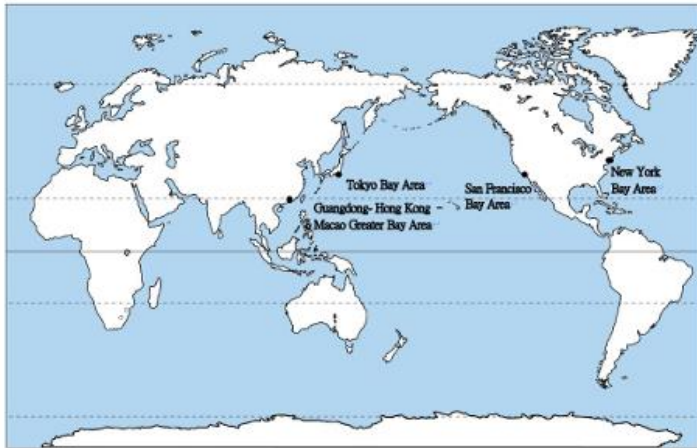
Comparison between the Guangdong-Hong Kong-Macao Greater Bay Area and other bay areas in the world

Edited by The Education University of Hong Kong

Personal, Social and Humanities Education Section, Curriculum Development Institute, Education Bureau

December 2021

Table 2: Land area and number of cities in the world's four bay areas in 2016

Bay area	Area (10,000 square kilometers)	Number of counties or cities
Guangdong-Hong Kong-Macao Greater Bay Area	5.65	11
New York Bay Area	2.15	25
San Francisco Bay Area	1.79	9
Tokyo Bay Area	1.36	4 (1 metropolis and 3 prefectures)

Source: Cushman & Wakefield (2018); Metropolitan Transportation Commission of San Francisco Bay Area (2017); Statistics Japan (Statistics Bureau Ministry of Internal Affairs and Communications) (2019)

Table 3: Population and percentage of national population of the world's four bay areas in 2016

Bay area	Population (ten thousands)	Percentage of national population (%)
Guangdong-Hong Kong-Macao Greater Bay Area	6,765	5
Tokyo Bay Area	4,383	28
New York Bay Area	2,370	7
San Francisco Bay Area	768	2

Source: Cushman & Wakefield (2018); Metropolitan Transportation Commission of San Francisco Bay Area (2017); Statistics Japan (Statistics Bureau Ministry of Internal Affairs and Communications) (2019)

Table 4: Gross domestic product (GDP) of the world's four bay areas in 2016

Bay area	GDP (trillion US$)	GDP per capita (US$)	GDP as a percentage of the country (%)
Tokyo Bay Area	1.8	41,070	41.0
New York Bay Area	1.4	69,307	4.4
Guangdong-Hong Kong-Macao Greater Bay Area	1.36	20,371	10.8
San Francisco Bay Area	0.76	99,802	7.7

Source: Cushman & Wakefield (2018); Metropolitan Transportation Commission of San Francisco Bay Area (2017); Statistics Japan (Statistics Bureau Ministry of Internal Affairs and Communications) (2019)

Table 5: Production and trade structure in the world's four bay areas in 2016

Bay area	Proportion of tertiary industry (2015)	Representative industry	Start-up industry	Direction of development
New York Bay Area	89.4%	Port trade	Port trade	World financial core
San Francisco Bay Area	82.8%	Trade, technological innovation	Trade technological innovation	Global high-technology research and development centre
Tokyo Bay Area	82.3%	Equipment manufacturing, steel, chemicals and logistics	Manufacturing innovation	Japan's core port industry
Guangdong-Hong Kong-Macao Greater Bay Area	62.2%	Finance, shipping, electronics and Internet	Foreign trade	Innovation and Technology

Hong Kong Policy Approach to Facilitating Data Flow and Data Security

Policy Statement on Facilitating Data Flow and Safeguarding Data Security in Hong Kong (Dec 2023) issued by Innovation, Technology and Industry Bureau (ITIB) and Digital Policy Office (formerly OGCIO) under "One Country, Two Systems"

- Background: PRC 14th Five-Year Plan for National Economic and Social Development of and the Long-Range Objectives Through the Year 2035 (Mar 13, 2021); Hong Kong I&T Development Blueprint (Dec 2022); ITIB and Cyberspace Administration of China signed "MOU on Facilitating Cross-boundary Data Flow Within the Guangdong-Hong Kong-Macao Greater Bay Area" (Jun 2023)
- 18 Action Items under 5 Broad Categories
  1. Advancing Digital Government and Enhancing Data Governance
  2. Formulating or Updating Policies, Guidelines and Laws
  3. Enhancing Cybersecurity Protection
  4. Bolstering the Digital Infrastructure
  5. Promoting Cross-boundary Data Flow

1. Advancing Digital Government and Enhancing Data Governance

Digital Policy Office (DPO) - merger of Office of Government Chief Information Office (OGCIO) and Efficiency Office in Jul 2024

- Ethical Artificial Intelligence Framework (Jul 2024 – updated from Aug 2023) for govt bureaux and depts implementing projects that involve use of AI technology
  - 12 Ethical AI Principles

Principle	Definition
Transparency and Interpretability	Organisations should be able to explain the decision-making processes of the AI applications to humans in a clear and comprehensible manner.
Reliability, Robustness and Security	Like other IT applications, AI applications should be developed such that they will operate reliably over long periods of time using the right models and datasets while ensuring they are both robust (i.e. providing consistent results and capable to handle errors) and remain secure against cyber-attacks as required by the relevant legal and industry frameworks.
Fairness	The recommendation/result from the AI applications should treat individuals within similar groups in a fair manner, without favouritism or discrimination and without causing or resulting in harm. This entails maintaining respect for the individuals behind the data and refraining from using datasets that contain discriminatory biases.
Diversity and Inclusion	Inclusion and diverse usership through the AI application should be promoted by understanding and respect the interests of all stakeholders impacted.
Human Oversight	The degree of human intervention required as part of AI application's decision-making or operations should be dictated by the level of the perceived severity of ethical issues.
Lawfulness and Compliance	Organisations responsible for an AI application should always act in accordance with the law and regulations and relevant regulatory regimes.

Principle	Definition
Data Privacy	Individuals should have the right to: (a) be informed of the purpose of collection and potential transferees of their personal data and that personal data shall only be collected for a lawful purpose, by using lawful and fair means, and that the amount of personal data collected should not be excessive in relation to the purpose. Please refer to the Data Protection Principles ("DPP")1 "Purpose and Manner of Collection" of the Personal Data (Privacy) Ordinance (the "PD(P)O")' (b) be assured that data users take all practicable steps to ensure that personal data is accurate and is not kept longer than is necessary. Please refer to the DPP2 "Accuracy and Duration of Retention" of the PD(P)O. (c) require that personal data shall only be used for the original purpose of collection and any directly related purposes. Otherwise, express and voluntary consent of the individuals is required. Please refer to the DPP3 "Use of Personal Data" of the PD(P)O. (d) be assured that data users take all practicable steps to protect the personal data they hold against unauthorised or accidental access, processing, erasure, loss or use. Please refer to the DPP4 "Security of Personal Data" of the PD(P)O. (e) be provided with information on (i) its policies and practices in relation to personal data, (ii) the kinds of personal data held, and (iii) the main purposes for which the personal data is to be used. Please refer to the DPP5 "Information to Be Generally Available" of the PD(P)O.
Safety	Throughout their operational lifetime, AI applications should not compromise the physical safety or mental integrity of mankind.
Accountability	Organisations are responsible for the moral implications of their use and misuse of AI applications. There should also be a clearly identifiable accountable party, be it an individual or an organisational entity (e.g. the AI solution provider).
Beneficial AI	The development of AI should promote the common good.
Cooperation and Openness	A culture of multi-stakeholder open cooperation in the AI ecosystem should be fostered.
Sustainability and Just Transition	The AI development should ensure that mitigation strategies are in place to manage any potential societal and environmental system impacts.

Table 1: Ethical AI Principles and Definition

1. Advancing Digital Government and Enhancing Data Governance

Digital Policy Office (DPO) Ethical Artificial Intelligence Framework (July 2024)
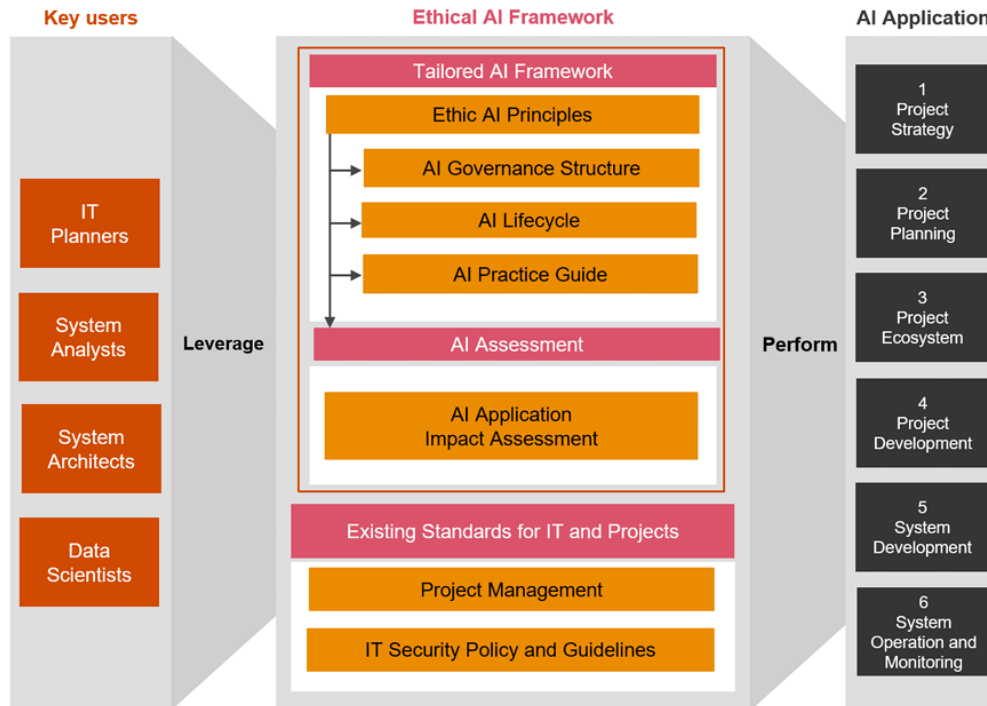

Figure 1: Overview of the Ethical AI Framework

AI Governance

AI governance refers to the practices and direction by which AI projects and applications are managed and controlled. The three lines of defence is a well-established governance concept in many organisations. Figure 2 shows the different defence lines and their roles.
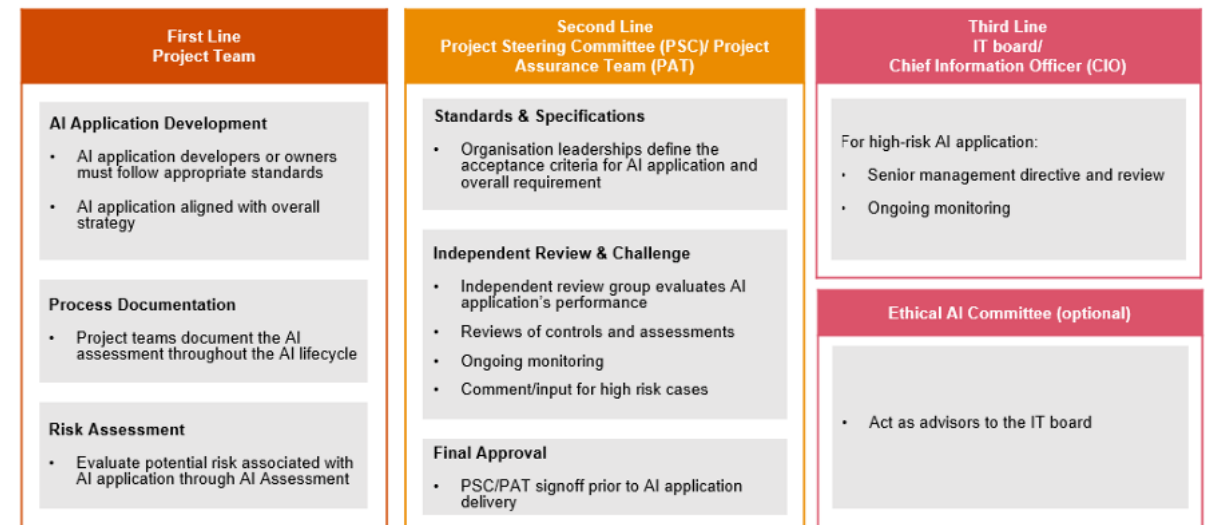

Figure 2: Lines of Defence Model

1. Advancing Digital Government and Enhancing Data Governance

Digital Policy Office (DPO) Ethical Artificial Intelligence Framework (July 2024)

AI Lifecycle

In order to structure the practices for organisations to follow when executing AI projects/creating AI applications, practices in different stages of the AI Lifecycle have been detailed in the AI Practice Guide (Please refer to Section 4 "AI Practice Guide" in the Ethical AI Framework for further details). A way to conceptualise the AI Lifecycle appears in the following 6-step schematic.
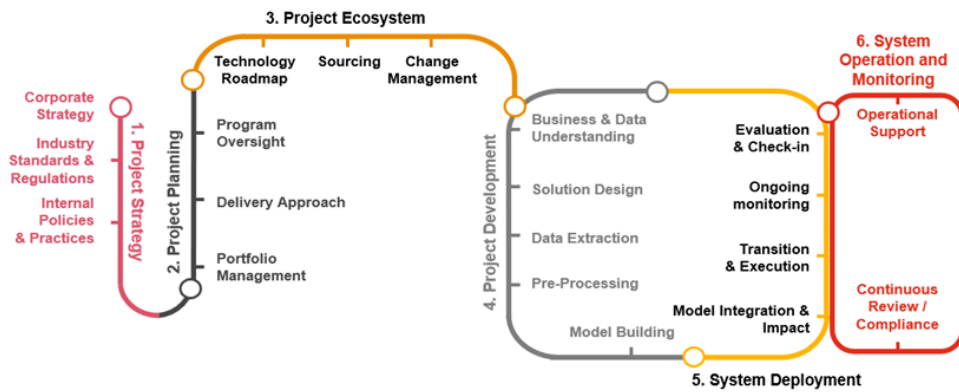
Figure 3: Overview of the AI Lifecycle

The AI Lifecycle is used to align the practices in the AI Practice Guide. The AI Lifecycle also aligns to a traditional System Development Lifecycle ("**SDLC**") model as depicted in Figure 4.
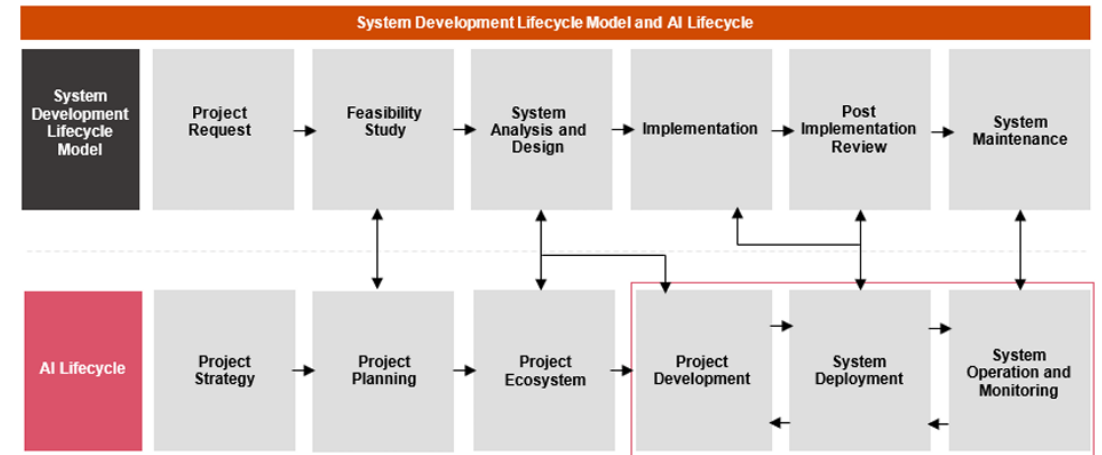
Figure 4: AI Lifecycle Aligned to a System Development Lifecycle Model

1. Advancing Digital Government and Enhancing Data Governance

Digital Policy Office (DPO) Ethical Artificial Intelligence Framework (July 2024)

AI Application Impact Assessment

The AI Application Impact Assessment should be conducted on an AI application at different stages of the AI Lifecycle. The AI Application Impact Assessment introduces a systematic thinking process for organisations to go through different aspects of considerations of individual applications for their associated benefits and risks whilst highlighting the need for additional governance activities and identifying follow-up actions to ensure necessary measures and controls required for implementing ethical AI.

The AI Application Impact Assessment template used for this assessment is in Microsoft Word format with sections for providing qualitative answers. Please refer to Appendix C "AI Application Impact Assessment Template" in the Ethical AI Framework document for details.

The AI Application Impact Assessment has the following components:

1. Risk Gating Criteria	2. AI Application Assessment Questions	3. AI Practice Guide Reference	4. Impact Considerations
• Determine the level of assessment required based on the AI application	• Consider the AI Lifecycle for the AI application • Each question has a core theme and is accompanied by explanatory text	• Some assessment questions reference the Practice Guide for additional guidance	• Identify beneficial and negative impacts to specific stakeholders and the controls required

Figure 5: AI Application Impact Assessment Components
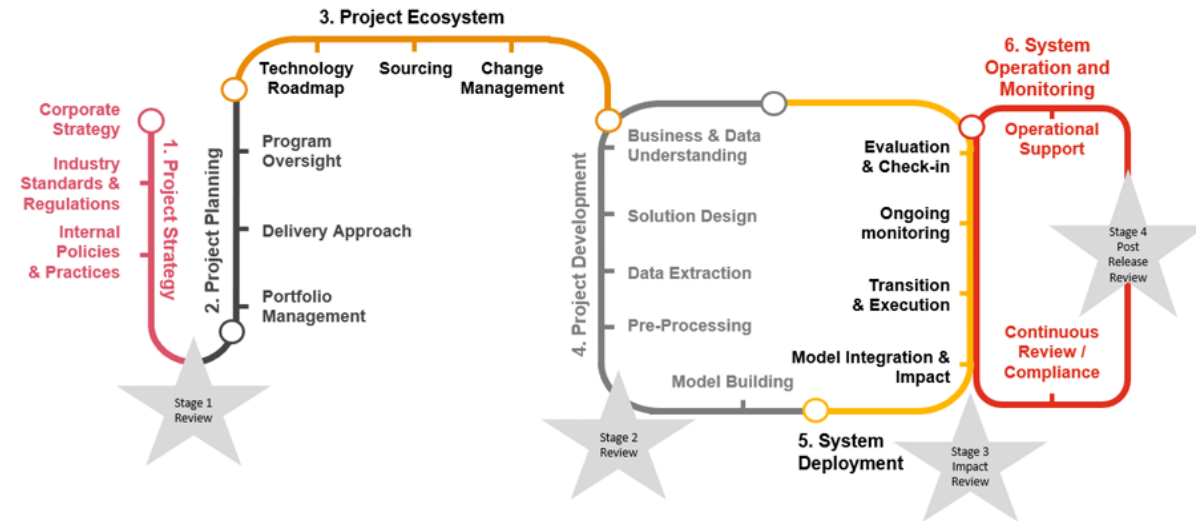
Figure 6: Stages for AI Application Impact Assessment

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- Personal Data (Privacy) Ordinance (PDPO) - when organisations develop and use AI process personal data, they would have to comply with the relevant requirements and six Data Protection Principles under the PDPO

Six Data Protection Principles

1. 收集目的及方式 COLLECTION PURPOSE & MEANS
2. 準確性、儲存及保留 ACCURACY & RETENTION
3. 使用 USE
4. 保安措施 SECURITY
5. 透明度 OPENNESS
6. 查閱及更正 DATA ACCESS & CORRECTION

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- "Guidance on the Ethical Development and use of Artificial Intelligence" (Aug 2021)

Designing your AI Solution Ethically

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- Findings of Compliance Checks on 28 Organisations (Feb 21, 2024)
  - 28 local orgs (telecom, finance, beauty, retail, transportation, education, govt depts)
  - 21 used AI in day-to-day operations (incl data analytics, HR, customer chatbots)
    - 19 established internal AI governance frameworks (eg, committee or officer)
    - 10 collected personal data through AI (all with PCPD statements)
    - 8 conducted privacy impact asssements
    - 10 implemented appropriate security measures
    - 9 retained personal data collected; 8 specified retention periods and delete or anonymize data when original purpose of collection achieved, remainder allowed data subjects to delete themselves

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)

- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)

Figure 1: *Data Stewardship Values and Ethical Principles for AI*

	Data Stewardship Values	Ethical Principles for AI
1	Being Respectful	• Accountability • Human Oversight • Transparency and Interpretability • Data Privacy
2	Being Beneficial	• Beneficial AI • Reliability, Robustness and Security
3	Being Fair	• Fairness

Figure 2: *Landscape of AI Developers, Vendors and Organisations Procuring / Implementing / Using AI*

Model Developers

Vendors (e.g., AI software / hardware companies, services integrators)

Organisations procuring, implementing and using AI (e.g., Government, public sector bodies or private sector companies)

Figure 3: *Model Personal Data Protection Framework*

Provide feedback for adjustment

ESTABLISH AI Strategy and Governance

CONDUCT Risk Assessment and Human Oversight

EXECUTE Customisation of AI Models and Implementation and Management of AI Systems

FOSTER Communication and Engagement with Stakeholders

Re-assess risks when there are significant changes

Fine-tune AI systems to address stakeholders' concerns

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)

Figure 4: Process of Procurement and Implementation of AI Models

1. Sourcing AI Solutions
2. Picking the Appropriate AI Solution
3. Collecting and Preparing Data
4. Customising AI Model for Particular Purpose
5. Testing, Evaluating and Validating AI Model
6. Testing and Auditing System and Components for Security and Privacy Risks
7. Integrating AI Solution into Organisation's System

Figure 5: Governance Considerations for Procuring AI Solutions

- Purpose(s) of Using AI
- Privacy and Security Obligations and Ethical Requirements
- International Technical and Governance Standards
- Criteria and Procedures for Reviewing AI Solutions
- Data Processor Agreements
- Policy on Handling Output Generated by the AI System
- Plan for Continuously Scrutinising Changing Landscape
- Plan for Monitoring, Managing and Maintaining AI Solution
- Evaluation of AI Suppliers

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
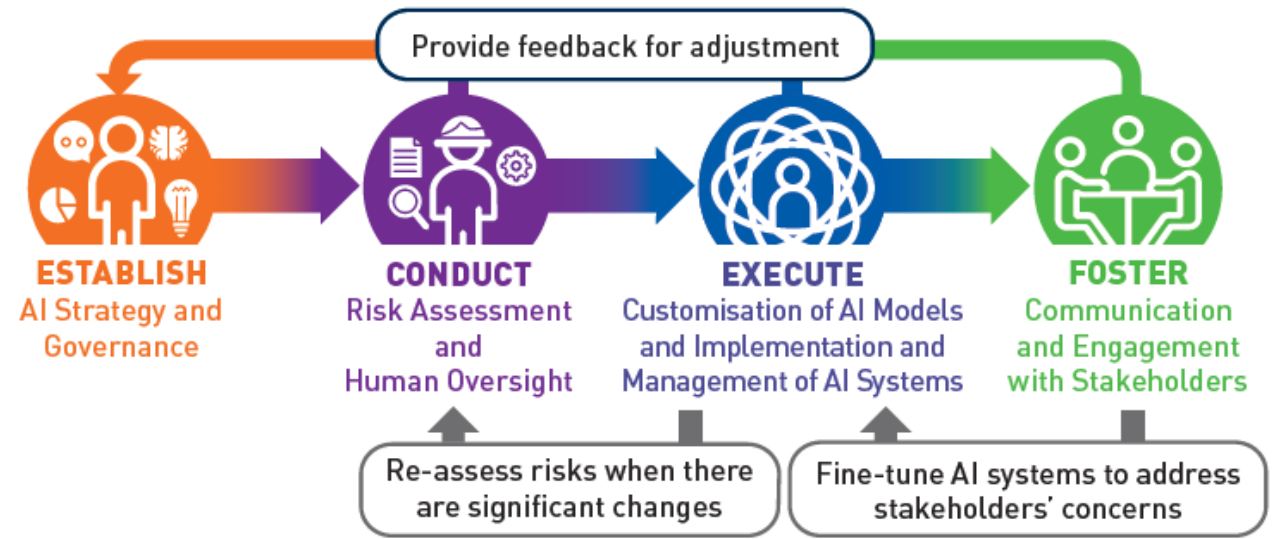- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)

AI Governance Committee

Participation by senior management and interdisciplinary collaboration should be the most significant attributes of an AI governance committee. A cross-functional team with a mix of skills and perspectives should be established, including business and operational personnel, procurement teams, system analysts, system architects, data scientists, cybersecurity professionals, legal and compliance professionals (including data protection officer(s)), internal audit personnel, human resources personnel and customer service personnel.

A C-level executive (such as a chief executive officer, chief information officer / chief technology officer, chief privacy officer or similar senior management position) should be designated to lead the cross-functional team.

(Optional) Independent AI and ethics advice may be sought from external experts. An additional ethical AI committee may be established to conduct an independent review when a project is sufficiently large, with a considerable impact and / or a high profile, and its ethical value may be challenged.

Examples of roles and responsibilities:

- Procurement teams should obtain AI solutions in accordance with the internal policies and procedures set out in the organisational AI strategy;
- System analysts, system architects and data scientists should focus on the customisation, implementation, monitoring and maintenance of AI solutions, and on the organisation's internal data governance processes;
- Legal and compliance professionals should focus on ensuring compliance with relevant laws and regulations (including data protection laws) as well as internal policies regarding the procurement, implementation and use of AI systems;
- Human reviewers should focus on reviewing the decisions and output of AI systems;
- Business and operational personnel should use AI in accordance with the policies and procedures of the organisations; and
- Customer service and public relations personnel should communicate with stakeholders, including customers, regulators and the general public, and address their concerns.

Figure 7: *Examples of Training*

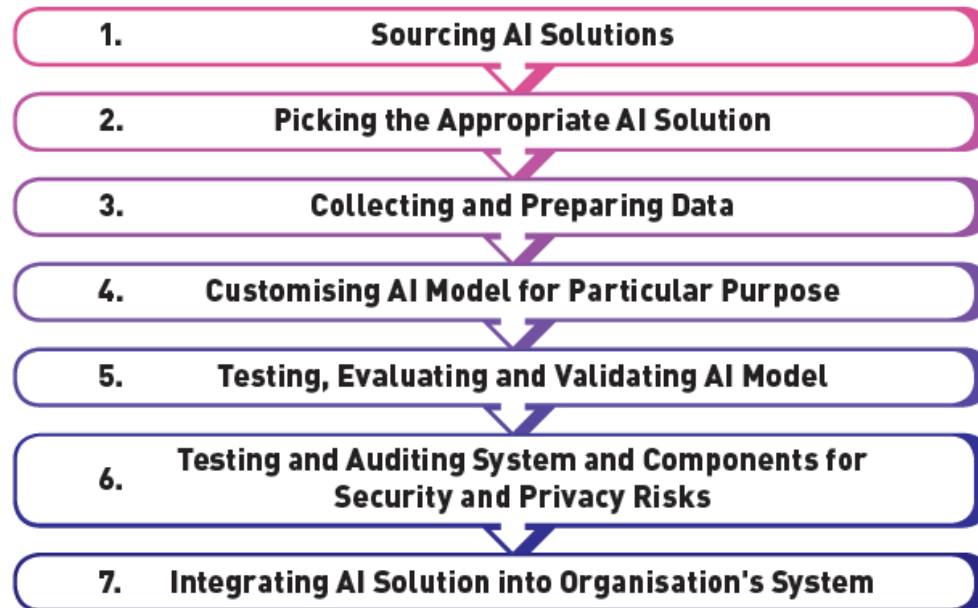| Recommended Personnel	Training Topics
System analysts / architects / data scientists	• Compliance with data protection laws, regulations and internal policies; cybersecurity risks
AI system users (including business and operational personnel)	• Compliance with data protection laws, regulations and internal policies; cybersecurity risks; general AI technology
Legal and compliance professionals	• General AI technology and governance
Procurement staff	• General AI technology and governance
Human reviewers	• Detection and rectification of any unjust bias, unlawful discrimination and errors / inaccuracies in the decisions made by AI systems or presented in the content
All staff performing work relating to AI system	• Benefits, risks, functions and limitations of the AI system(s) used by the organisation

Designing your AI Solution Ethically

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)

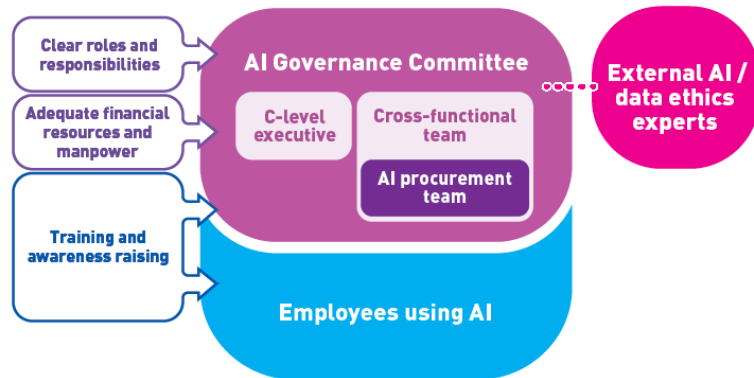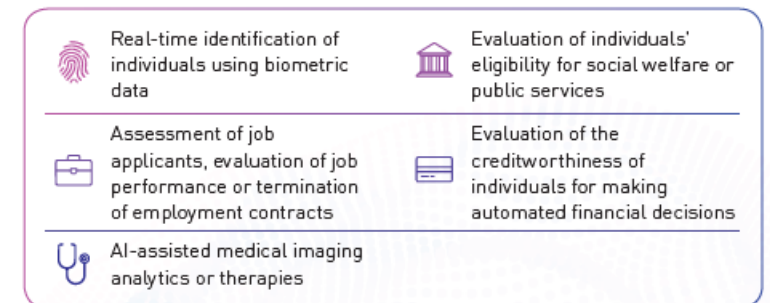Figure 8: Governance Structure

Figure 9: Process of Risk Assessment

Figure 11: Risk-based Approach to Human Oversight

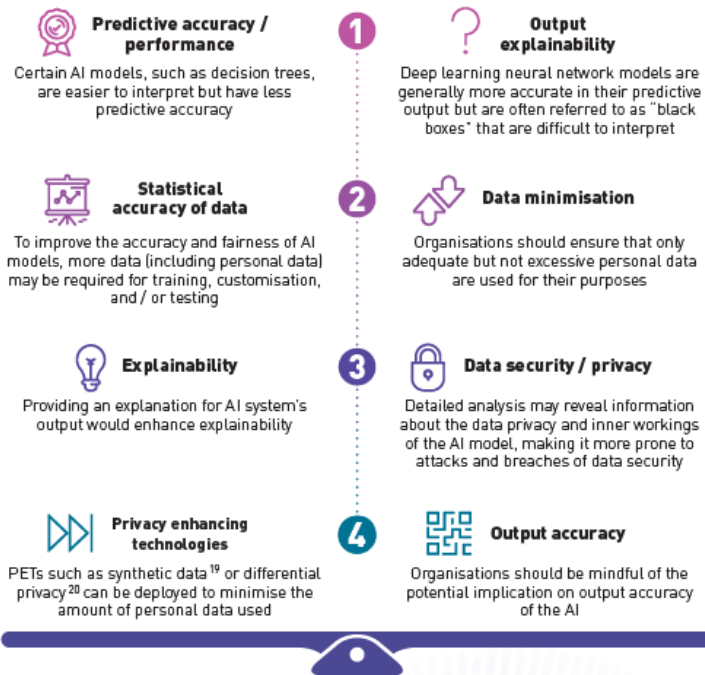Figure 12: Examples of AI Use Cases that May Incur Higher Risk

Designing your AI Solution Ethically

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)
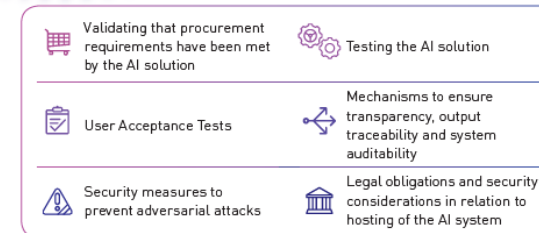
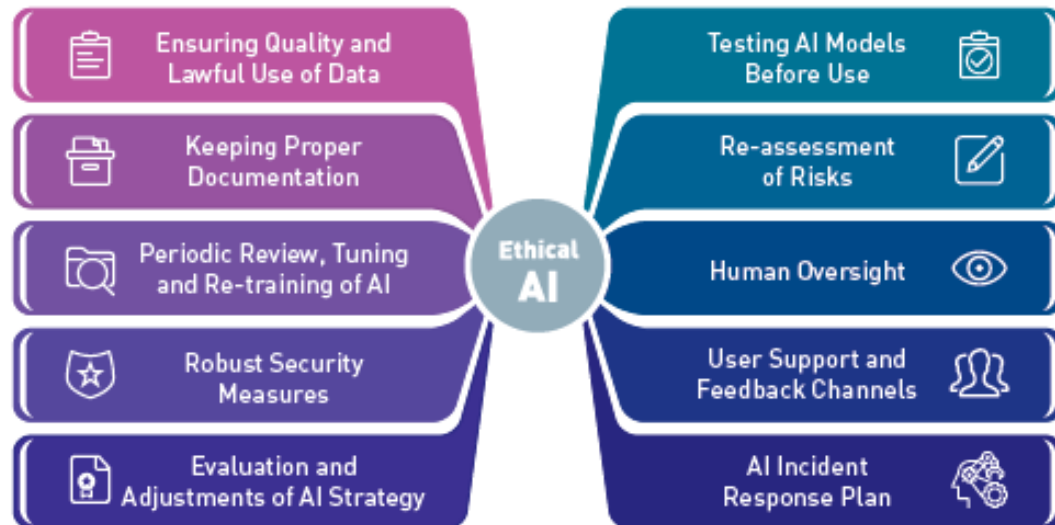**Figure 13:** Examples of Risk Mitigation Trade-offs

Predictive accuracy / performance
Certain AI models, such as decision trees, are easier to interpret but have less predictive accuracy

① Output explainability
Deep learning neural network models are generally more accurate in their predictive output but are often referred to as "black boxes" that are difficult to interpret

Statistical accuracy of data
To improve the accuracy and fairness of AI models, more data (including personal data) may be required for training, customisation, and / or testing

② Data minimisation
Organisations should ensure that only adequate but not excessive personal data are used for their purposes

Explainability
Providing an explanation for AI system's output would enhance explainability

③ Data security / privacy
Detailed analysis may reveal information about the data privacy and inner workings of the AI model, making it more prone to attacks and breaches of data security

Privacy enhancing technologies
PETs such as synthetic data [19] or differential privacy [20] can be deployed to minimise the amount of personal data used

④ Output accuracy
Organisations should be mindful of the potential implication on output accuracy of the AI

Figure 14: Major Customisation and Management Processes

Data Preparation and Management → Customisation and Implementation → Management and Continuous Monitoring

Figure 15: 4 Aspects of Data Preparation

Data Preparation
- Compliance with requirements of the PDPO
- Minimisation of personal data involved in customisation and use of AI
- Management of data for customising and using AI
- Proper documentation of handling of data

Figure 17: Examples of AI Solution Implementation Considerations
- Validating that procurement requirements have been met by the AI solution
- Testing the AI solution
- User Acceptance Tests
- Mechanisms to ensure transparency, output traceability and system auditability
- Security measures to prevent adversarial attacks
- Legal obligations and security considerations in relation to hosting of the AI system

Figure 18: AI Incident Response Plan

Defining an AI Incident
- Organisations should devise a definition in the context of their AI systems.
- An AI incident may be defined as "an event where the development or use of an AI system [allegedly] caused harm to person(s), property, or the environment, including by infringing upon human rights, such as privacy and non-discrimination; [where the] harm involves bodily injury or death, it could be considered to be a 'serious incident'". [47].

Monitoring for AI Incidents
- Closely tied to the risk assessment process, categories of foreseeable harms should be noted and monitored, and procedures for addressing unforeseeable harms that emerge should be devised.
- Organisations may note the past AI incidents that are documented in the "AI Incident Database" [48].

Reporting an AI Incident
- Internal policies and procedures should be established to enable employees to flag incidents, and to enable other stakeholders (i.e., business partners, customers) to report any incidents through feedback channels.

Containing an AI Incident
- Personnel should be designated as responsible for pressing the "pause" or "stop" button on the AI system according to established policies and procedures to disconnect the systems affected from other operating systems.
- Relevant regulatory authorities and any impacted individuals should be informed as soon as practicable.

Investigating an AI Incident
- Relevant personnel (including those responsible for implementing the AI system) should conduct a thorough review and investigation and apply technical fixes.
- Result of the investigation should be reported in line with the organisation's AI policies.
- The AI system should only resume operation when it has been confirmed that the risk of further harm or unintended consequences is minimised.

Recovering from an AI Incident
- Salient findings from the incident investigation should be documented.
- Findings may necessitate the revision of internal policies and procedures for procurement, changes in the implementation and use of AI in the organisation's AI strategy and updates to internal training.

1. Advancing Digital Government and Enhancing Data Governance

Hong Kong Privacy Commissioner for Personal Data (PCPD)
- "Artificial Intelligence: Model Personal Data Protection Framework" (Jun 11, 2024)

Figure 19: Management of AI Systems

Figure 20: Communication and Engagement with Stakeholders

2. Formulating or Updating Policies, Guidelines and Laws

- Secretary for Innovation, Technology & Industry Sun Dong response to question at Legislative Council (Jan 24, 2024)
  "Government has commissioned the InnoHK research centre specialised in generative AI to study and suggest appropriate rules and guidelines on the accuracy, responsibility and information security in the technology and application of generative AI technologies. We will study the appropriate strategy and measures with reference to suggestions by industry experts, with a view to balancing the need to develop AI technology and safeguarding security, etc."

- "ITIB and DPO will review the existing arrangements on the collection, use, processing, protection and sharing of data, and follow up on feasible measures to tackle pain points in specific areas."

- "Constitutional and Mainland Affairs Bureau will study possible amendments to the PDPO to align with the latest international developments in privacy protection, strengthen personal data protection, and address the challenges posed by cyber technologies."

2. Formulating or Updating Policies, Guidelines and Laws

- Current approach is not wholesale AI regulation (notwithstanding Cyberspace Administration of China's Interim Measures for the Administration of Generative Artificial Intelligence Services effective Aug 2023): per Secretary for Innovation, Technology and Industry response to LegCo question (May 21, 2023):

  - Eg, Crimes (Amendment) Ordinance 2021 introduced the offences of publication or threatened publication of intimate images without consent. The "anti-voyeurism" offence is also applicable to intimate images that have been altered (including that altered by AI technology)

2. Formulating or Updating Policies, Guidelines and Laws

- Commerce and Economic Development Bureau (CEDB) and Intellectual Property Department (IPD) conducted Public Consultation on amendment to Copyright Ordinance to protect AI development (Jul 2024)
  (a) Copyright protection of AI-generated works (Data-mining exemption like Singapore, Japan)
  (b) Copyright infringement liability for AI-generated works
  (c) Possible introduction of specific copyright exception
  (d) Other issues relating to generative AI (eg, Deepfakes; AI Transparency)

| LDMA works	Originality requirement	Creator in real life	Author[11]	First copyright owner[12]	Duration of copyright[13]	Moral rights[14]
Ordinary LDMA works	Yes	Human author	Human author		Author's life plus 50 years after death	• right to be identified as the author • right to object to derogatory treatment of the work • right against false attribution of a work
CG LDMA works	Yes	Computer (without human author)	Person by whom the arrangements necessary for the creation of the work are undertaken[15]		50 years from which the work was made	• right against false attribution of a work

Non-LDMA works	Originality requirement	Creator in real life	Author[17]	First copyright owner[18]	Duration of copyright[19]	Moral Rights[20]
Sound recordings				Producer[21]	50 years from which the recording was made/released	N.A.
Films	No	No statutory restrictions/ requirements in the CO that exclude computer as a creator		Producer[21] and human principal director	i) 50 years from which the death occurs of the last to die of the following persons– • principal director • author of screenplay • author of dialogue • composer of music specially created for and used in the film ii) in case there is no such person falling in (i) above, 50 years from which the film was made	• right to be identified as the director • right to object to derogatory treatment of the work • right against false attribution of a work
Broadcasts				Person making the broadcast	50 years from which the broadcast was made	N.A.
Cable programmes				Person providing the cable programme service	50 years from which the programme was included in the cable programme service	N.A.
Typographical arrangement of published editions				Publisher	25 years from which the edition was first published	N.A.

3. Enhancing Cybersecurity Protection

Background:

- **PRC Standing Committee of the National People's Congress** passed Data Security Law requiring system for data protection, risk assessment, reporting, monitoring, early warnings and data security emergency response system (Jun 20, 2021)
- **Security Regulations**, authorised by Security Bureau, provides directives on what documents, material and information need to be classified and to ensure given an adequate level of protection in relation to the conduct of government business.
- **Government IT Security Policy and Guidelines**, established by DPO, aim to facilitate implementation of information security measures to safeguard information assets (eg, Info security management systems (ISO/IEC 27001: 2022) and Info security, cybersecurity & privacy protection (ISO/IEC 27002: 2022))
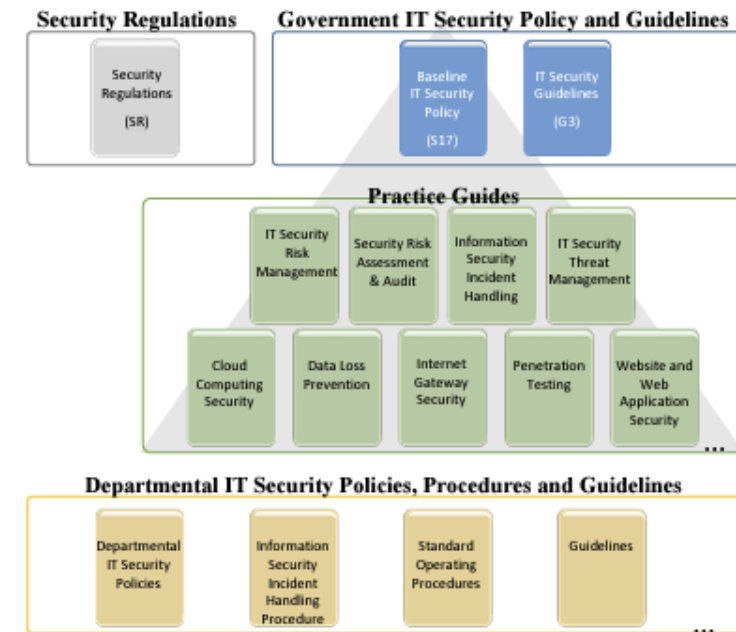
3. Enhancing Cybersecurity Protection

Hong Kong Police Force (HKPF) established
- Critical Infrastructure Security Coordination Centre (CISCC), which sought to strengthen self-protection and self-restoration capabilities of these infrastructures through public-private co-operation, risk management, on-site security inspections, promotion of restoration plans and security designs.

- Cyber Security Centre (CSC) under the HKPF's Cyber Security and Technology Crime Bureau provides support to critical infrastructures by conducting timely cyber threat audits and analyses to prevent and detect cyber attacks against them.

DPO (formerly OGCIO)
- Reached consensus with Bureau of Cyber Security of Cyberspace Administration of China on co-operation in 2016 to strengthen co-ordination and promote exchanges and co-operation in cyber security between Mainland China and Hong Kong.
- Works with the National Computer Network Emergency Response Technical Team/Coordination Center of China to obtain related cyber security vulnerability information in a timely manner through the China National Vulnerability Database and arrange preventive measures.

The University of Hong Kong
LiTE LAB
Law | Innovation | Technology | Entrepreneurship

3. Enhancing Cybersecurity Protection

Public consultation on **Protection of Critical Infrastructure (Computer System) Bill** on legislative framework to regulate cybersecurity obligations of critical infrastructure operators (CIO) prepared jointly by Security Bureau, DPO and HKPF (July 2, 2024)

- **Background:** Mainland China Cybersecurity Law 2016 and Regulation for Safe Protection of Critical Information Infrastructure 2021; Hong Kong Law Reform Commission (HKLRC) separately released Consultation Paper proposing the New Cybercrime Offences (July 2022)
- **New Commissioner's Office under Security Bureau**: investigative powers and designate industry-specific regulators of essential services sectors (eg, HKMA and Communications Authority) to monitor compliance.
- **Obligations**: keep Office updated on CI ownership and operatorship, detailed plans, risk assessment, emergency response plan, report security incidents
- **Critics**: AmChamHK suggested limit to CIOs located within Hong Kong; Bloomberg article noted that US firms had expressed concerns, including that it could grant Hong Kong govt "unusual access to their computer system."; government rebuke

4. Bolstering the Digital Infrastructure

- **InnoHK from HK$10 billion allocation in 2018**
  - To transform the city into a global innovation powerhouse through creating research clusters in collaboration with leading institutions, including AIR@InnoHK at Hong Kong Science & Technology Park focusing on AI and robotics technologies.
  - As of Apr 30, 2024, there are a total of 29 centres with a new InnoHK centre specialising in R&D of GenAI technology established in Sep 2023

- **Government's Consented Data Exchange Gateway (CDEG)**
  - To enable citizens to authorise govt depts to use their personal information stored in other depts
  - HKMA's Commercial Data Interchange (CDI) and the Government's CDEG connection announced open to all CDI participants (Aug 26, 2024)

4. Bolstering the Digital Infrastructure

Item	($ million)			
	2024-25	2025-26	2026-27	Total
Computing power subsidy	857	857	1,142	2,856
Enhancing cyber and data security	42	32	26	100
Promoting the development of AI ecosystem	6	16	22	44
Total	905	905	1,190	3,000

- HK$ 3 billion AI Subsidy Scheme announced
  - To support local universities, research institutes and enterprises in harnessing the computing power of the AI Supercomputing Center at Cyberport

- CEDB to continue to enhance coverage and capacity of 5G network & infrastructure
  - Amending Telecommunications Ordinance and revising relevant guidelines to ensure appropriate space made available in new buildings for installation of telecommunications facilities by mobile network operators, as well as amending Inland Revenue Ordinance to provide tax incentives to operators in respect of spectrum utilization fees to promote development of 5G infrastructure

4. Bolstering the Digital Infrastructure

- **Government's iAMSmart mobile app** as single portal for online government services
    - Seeking to increase the utilization rate of "iAM Smart"
    - Add more features – eg, facial recognition authentication function (with consent)
    - Cross-boundary e-Government services - OGCIO is actively exploring with Government Services and Data Management Bureau of Guangdong Province as one of the means for authentication on the "Unified Identity Authentication Platform of Guangdong Province"
    - Developing business version of "iAM Smart" "Digital Corporate Identity" (CorpID)

5. Promoting Cross-boundary Data Flow

ITIB and Cyberspace Administration of China signed MOU on Facilitating Cross-boundary Data Flow Within the Guangdong-Hong Kong-Macao Greater Bay Area (Jun 2023)

- Seek to implement in an orderly manner measures to facilitate crossboundary data flow in the GBA, including the early and pilot implementation of the arrangement to streamline and facilitate the compliance procedures for the flow of personal information relating to banking, credit checking and healthcare from GBA to Hong Kong

Facilitation measure on Standard Contract for Cross-boundary Flow of Personal Information within Guangdong-Hong Kong-Macao Greater Bay Area (Dec 2023)

- Voluntary GBA Standard Contract to outline obligations and responsibilities of personal information processor and recipient registered (applicable to organisations)/located (applicable to individuals) in 9 Mainland cities within GBA (ie, Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen and Zhaoqing), and to conduct cross-boundary flow of personal information between these Mainland cities and Hong Kong (i.e., flow of the personal information from the 9 Mainland cities in the GBA to Hong Kong, and vice versa).
- Administer by DPO and Cyberspace Administration of Guangdong Province

Hong Kong AI regulation – sectoral approach in finance

- HKMA High-Level Principles on AI (Nov 1, 2019)
  - Board and senior management accountable for the outcome of AI applications
  - Possessing sufficient expertise
  - Ensuring an appropriate level of explainability of AI applications
  - Using data of good quality
  - Conducting rigorous model validation
  - Ensuring auditability of AI applications
  - Implementing effective management oversight of third-party vendors
  - Being ethical, fair and transparent
  - Conducting periodic reviews and on-going monitoring
  - Complying with data protection requirements
  - Implementing effective cybersecurity measures
  - Risk mitigation and contingency plan

Designing your AI Solution Ethically

Hong Kong AI regulation – sectoral approach in finance

- HKMA and Cyberport launch GenAI Sandbox (Aug 13, 2024)
  - HKMA Invitations extended to banks (Sep 20, 2024)
- HKMA Consumer Protection in respect of use of GenAI (Aug 19, 2024)
  - Governance and accountability
  - Fairness
  - Transparency and disclosure
  - Data privacy and protection
  - Proactive use of BDAI and GenAI in enhancing consumer protection
- HKMA Use of AI for Monitoring of Suspicious Activities (Sep 9, 2024)
- Research Paper of Generative AI in the Financial Services Sector (Sep 27, 2024)
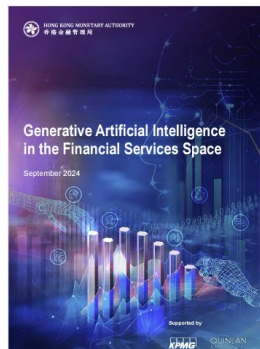
Hong Kong AI regulation – sectoral approach in finance

Figure 2: A.I. Supporting Pillars
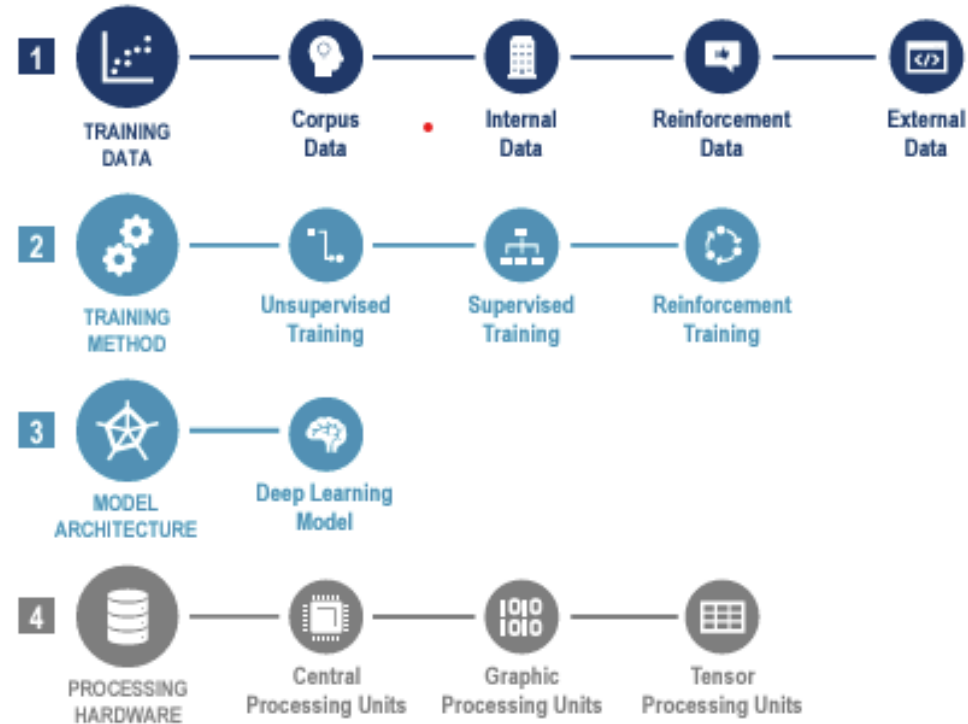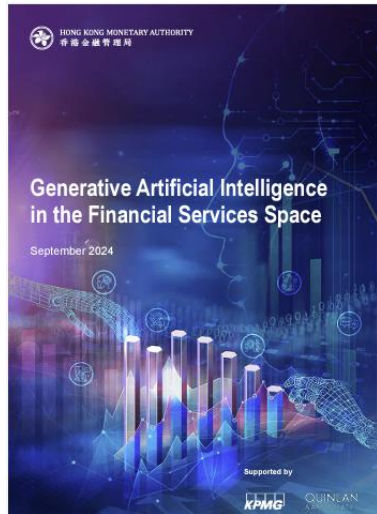
Source: Subject matter expert interviews, Quinlan & Associates analysis

Hong Kong AI regulation – sectoral approach in finance

Figure 14: GenA.I. Workflow Applicability
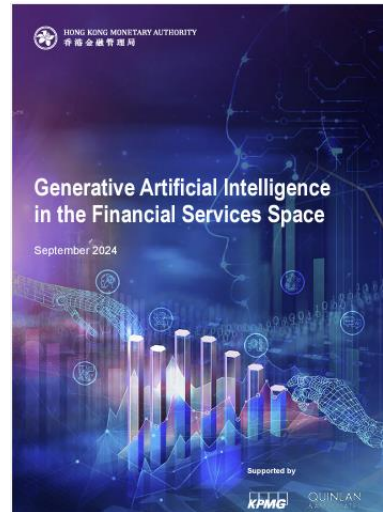
	Description	Human-in-the-Loop
IDENTIFY NEEDS	Evaluate key objectives, such as organisational goals, regulatory requirements, and customer demands, to pinpoint challenges or opportunities that necessitate actions.	MODERATE Augmentation
GENERATE IDEAS	Brainstorm and develop potential strategies and approaches to effectively address the identified challenges or opportunities.	MODERATE Augmentation
COLLECT DATA	Gather pertinent data (e.g., market reports, customer information, etc.) from both internal and external sources, to support informed decision-making.	LOW Automation
ANALYSE DATA	Apply analytical tools and techniques to process and interpret collected data, identifying significant patterns, trends, and anomalies.	LOW Automation
DERIVE INSIGHTS	Convert data analysis into actionable insights with direct implications for decision-making.	LOW Automation
GENERATE OUTPUT	Create outputs based on findings that align with and effectively address the defined challenges or opportunities.	LOW Automation
REVIEW OUTPUT	Rigorously review outputs for content accuracy, tone consistency with organisational standards, and alignment to industry best practices.	HIGH GenA.I. Not Applicable
TAKE ACTION	Implement decisions (e.g., trade execution, client request handling, etc.) based on the reviewed outputs.	HIGH GenA.I. Not Applicable

Source: Quinlan & Associates analysis

Designing your AI Solution Ethically

Hong Kong AI regulation – sectoral approach in finance

Figure 21: Key Regulatory Principles of A.I.

COMMON PRINCIPLES

Governance & Accountability
Assign clear stakeholder responsibilities in the governance structure and ensure an appropriate level of explainability

Fairness
Ensure equal and inclusive treatment of all users, prevent bias, and avoid discrimination against any certain group

Data Privacy & Protection
Implement robust security measures on the system, seek data privacy consent, and communicate the intended usage of data

Transparency & Disclosure
Provide users with clear and sufficient disclosures about the use of A.I. in the service provided to protect their rights

ADDITIONAL PRINCIPLES

Reliability
Operate the A.I. system with resilience, using high-quality data to deliver accurate and reliable responses

Sustainability
Actively monitor and manage A.I.'s environmental and social impacts to ensure sustainable development

Source: Regulatory disclosures, Quinlan & Associates analysis

Generative Artificial Intelligence in the Financial Services Space

September 2024

Designing your AI Solution Ethically

Hong Kong AI regulation – sectoral approach in finance

Figure 26: GenA.I. Adoption Considerations

Source: Quinlan & Associates analysis

Hong Kong AI regulation – sectoral approach in finance

Figure 27: GenA.I. Governance Structure

REGULATORY PRINCIPLES

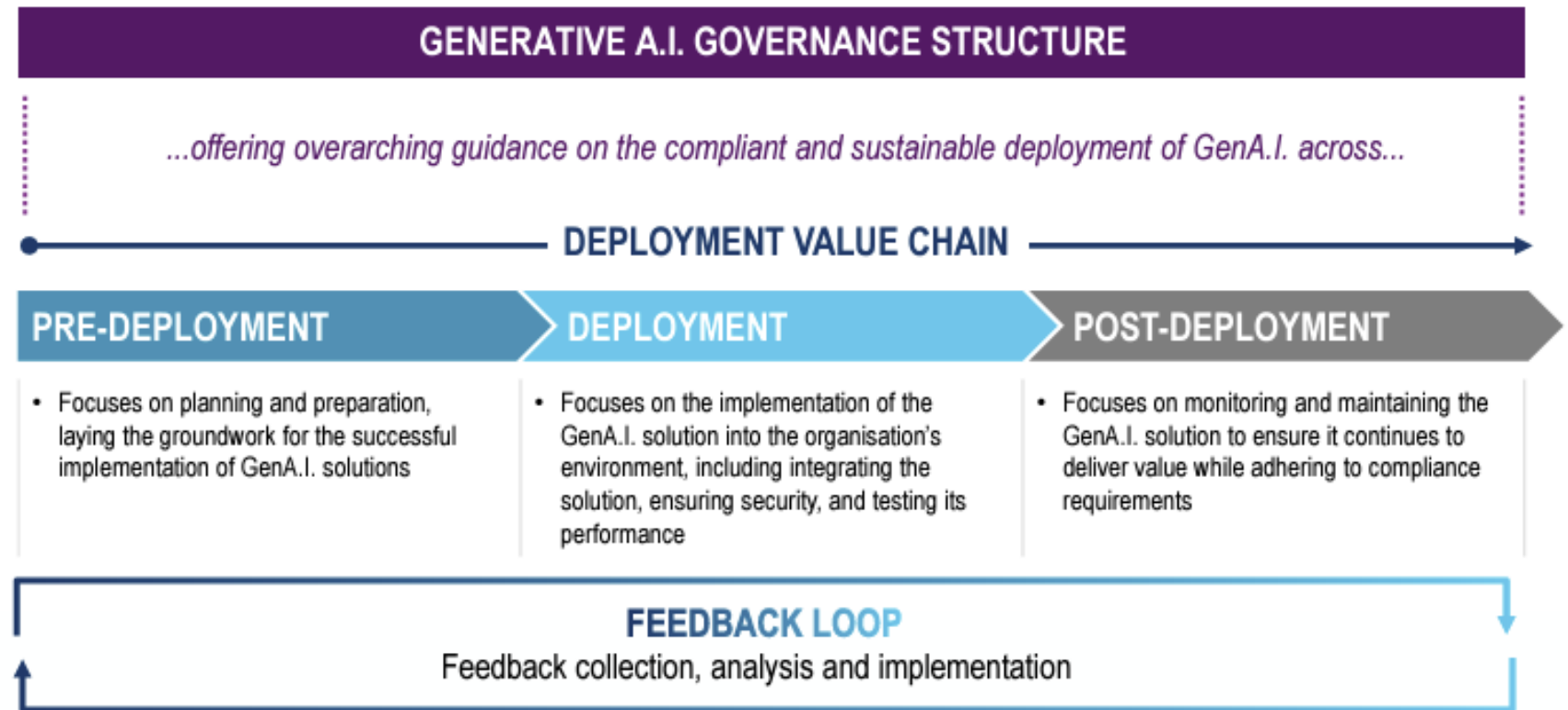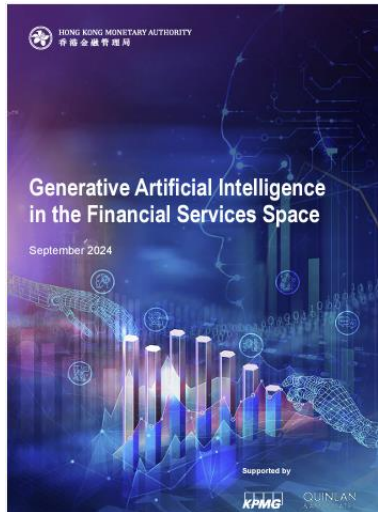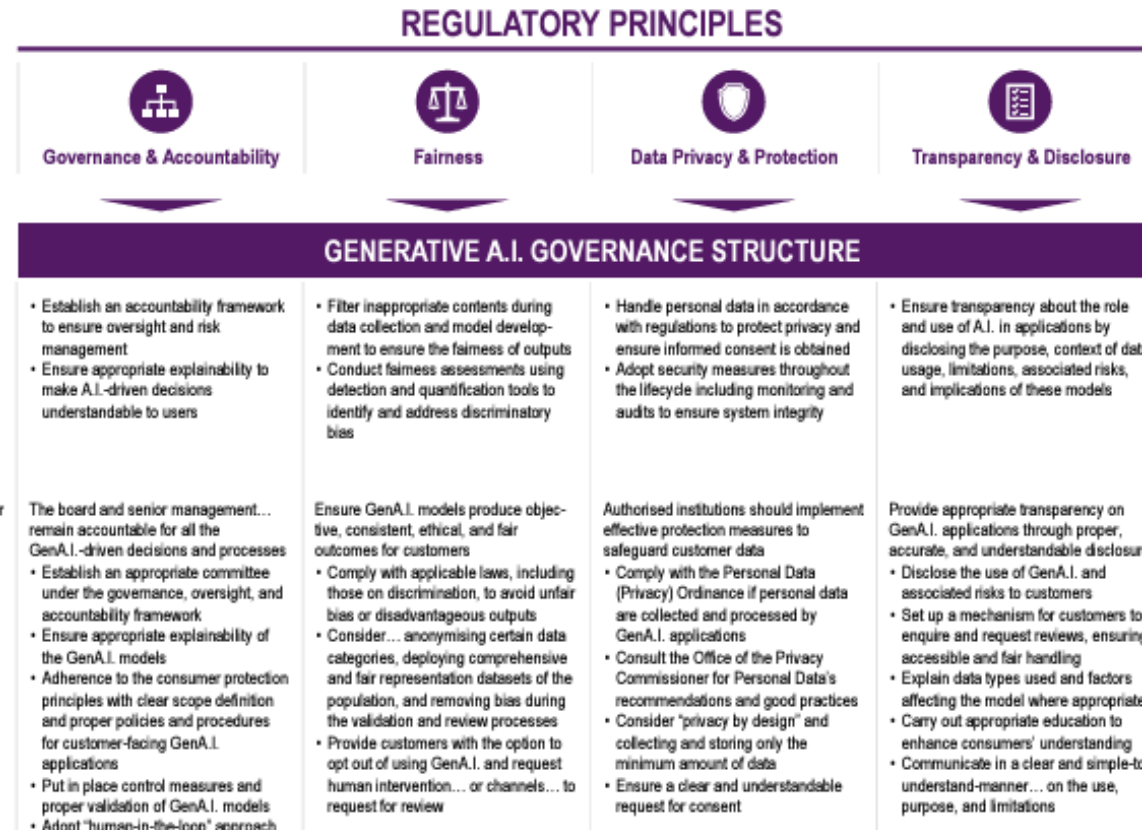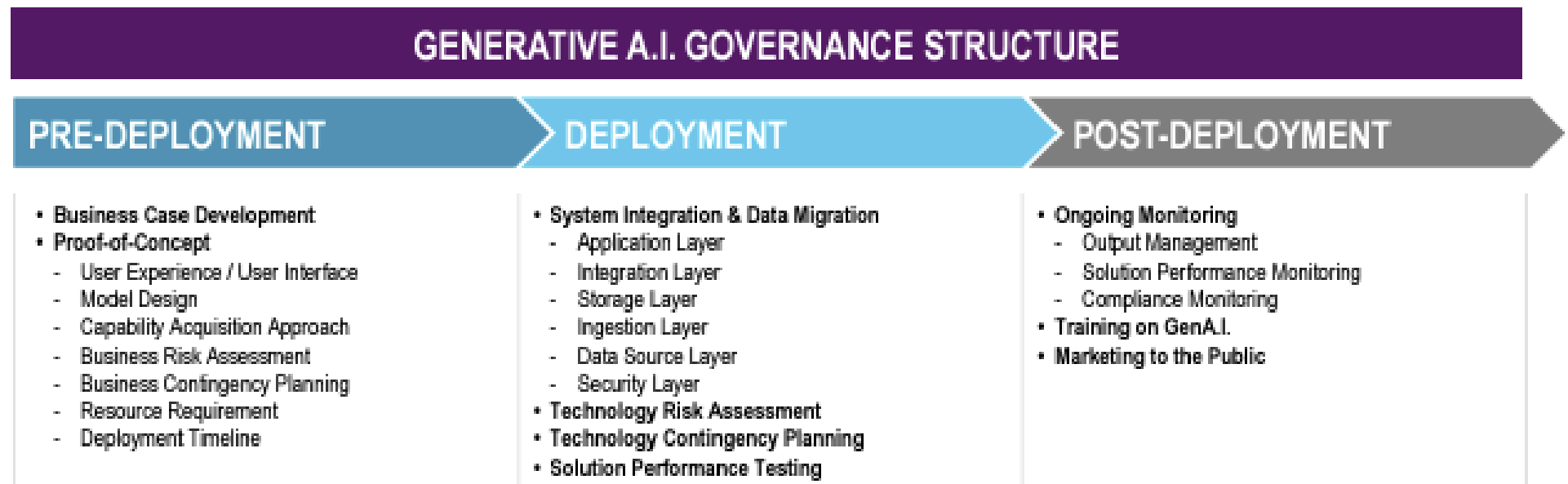|	Governance & Accountability	Fairness	Data Privacy & Protection	Transparency & Disclosure
GENERATIVE A.I. GOVERNANCE STRUCTURE				
Global Standards	• Establish an accountability framework to ensure oversight and risk management • Ensure appropriate explainability to make A.I.-driven decisions understandable to users	• Filter inappropriate contents during data collection and model development to ensure the fairness of outputs • Conduct fairness assessments using detection and quantification tools to identify and address discriminatory bias	• Handle personal data in accordance with regulations to protect privacy and ensure informed consent is obtained • Adopt security measures throughout the lifecycle including monitoring and audits to ensure system integrity	• Ensure transparency about the role and use of A.I. in applications by disclosing the purpose, context of data usage, limitations, associated risks, and implications of these models
HKMA Circular	The board and senior management... remain accountable for all the GenA.I.-driven decisions and processes • Establish an appropriate committee under the governance, oversight, and accountability framework • Ensure appropriate explainability of the GenA.I. models • Adherence to the consumer protection principles with clear scope definition and proper policies and procedures for customer-facing GenA.I. applications • Put in place control measures and proper validation of GenA.I. models • Adopt "human-in-the-loop" approach	Ensure GenA.I. models produce objective, consistent, ethical, and fair outcomes for customers • Comply with applicable laws, including those on discrimination, to avoid unfair bias or disadvantageous outputs • Consider... anonymising certain data categories, deploying comprehensive and fair representation datasets of the population, and removing bias during the validation and review processes • Provide customers with the option to opt out of using GenA.I. and request human intervention... or channels... to request for review	Authorised institutions should implement effective protection measures to safeguard customer data • Comply with the Personal Data (Privacy) Ordinance if personal data are collected and processed by GenA.I. applications • Consult the Office of the Privacy Commissioner for Personal Data's recommendations and good practices • Consider "privacy by design" and collecting and storing only the minimum amount of data • Ensure a clear and understandable request for consent	Provide appropriate transparency on GenA.I. applications through proper, accurate, and understandable disclosure • Disclose the use of GenA.I. and associated risks to customers • Set up a mechanism for customers to enquire and request reviews, ensuring accessible and fair handling • Explain data types used and factors affecting the model where appropriate • Carry out appropriate education to enhance consumers' understanding • Communicate in a clear and simple-to-understand-manner... on the use, purpose, and limitations

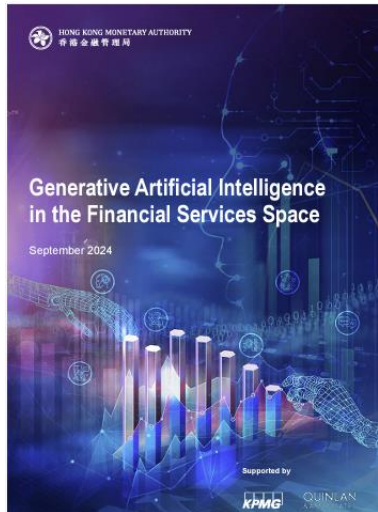Source: Regulatory disclosures, Quinlan & Associates analysis

Designing your AI Solution Ethically

Hong Kong AI regulation – sectoral approach in finance

Figure 28: Adoption Considerations Across the Value Chain

Source: Quinlan & Associates analysis

Hong Kong AI regulation – sectoral approach in finance

- [HK Government to consider rules for AI use in finance](#) (Standard, Sep 17, 2024)
  - "Local regulators are also trying to resolve some of the confusion around AI in Hong Kong, a city that's caught up in the US-Chinese technology conflict. Many consumers and corporations can't easily access some of the hottest services from OpenAI's ChatGPT to Google's Gemini, because US tech leaders are likely nervous about running afoul of the Chinese territory's rules, analysts say. Officials expect to unveil their statement around late October during Fintech Week, one of the industry's most important annual gatherings."

International efforts to coordinate from Hong Kong

1. Hong Kong SAR as part of "One Country Two Systems"
   - "Microsoft maintains AI services in Hong Kong, as OpenAI curbs API access from China" (SCMP, Jun 27, 2024)
   - "As OpenAI blocks China, developers scramble to keep GPT access through VPNs" (SCMP, Jul 10, 2024)
   - "China's Views on AI Safety Are Changing—Quickly" (Carnegie Endowment for International Peace, Aug 27, 2024)

International efforts to coordinate from Hong Kong

2. Privacy regulator conferences and forums
   - Eg, Asia-Pacific Privacy Authorities (APPA) Forum
   - Global Privacy Assembly – Working Group on Ethics and Data Protection in Artificial Intelligence
3. University initiatives
   - International AI Cooperation and Governance Forum 2023 (HKUST and Tsinghua U) (Dec 8-9, 2023)

Designing your AI Solution Ethically

Chiron HWITL framework survive introduction of GenAI ?

Chiron, the "justest and wisest of centaurs"

© Brian Tang

HKU LiTE Machine Lawyering Conference 2021

Chiron Imperative: An AI Economy "Humans-Working-In-The-Loop" Framework for Wise and Just AI-Human Centaurs

Deconstruction for Holistic Approach of AI Economy "Humans-Working-In-The-Loop" for better policy and regulatory recommendations regarding responsibility, accountability and liability as AI-Human Centaurs

Human as AI Data Preparer → Human as AI User-Trainer → Human as AI Governance → Human as AI Explainer or Interpreter → Human as AI Creator → Human as Customer-User of AI Products and Services

This paper: Human as AI Trainer / AI Data Preparer
a) Roles, tasks and titles
b) Why Matters

"Garbage In, Garbage Out"
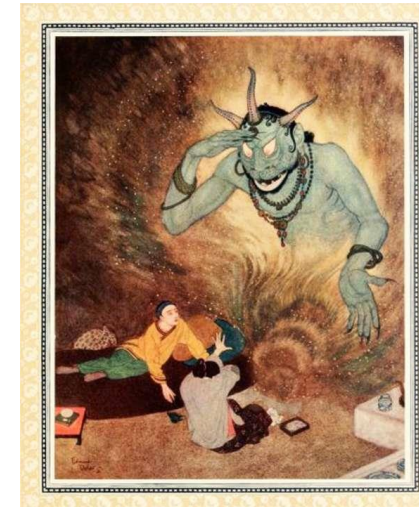
Cheeky Djinn, or "how do you make your wish?"
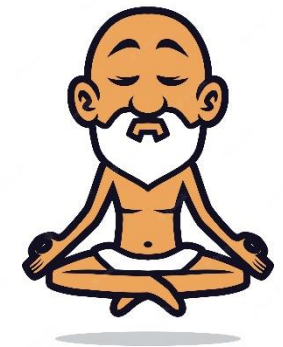
Stochastic Guru, or "seeking of insights (to follow)"

bwtang@hku.hk

Chiron HWITL GenAI Roles: Market-based Mechanisms for GenAI externalities?

GenAI Consumers: Universities
Law Firms/ Legal Departments
Students; Lawyers; Clients

Human-Working-In-The-Loop Role	Concerns/ Externalities
Human as AI Trainer/ Data Preparer	Bias training data; Copyright infringement; Data worker pay & conditions
Human as AI User Trainer	Confidentiality/privacy of prompting
Human as AI Governance	Credibility of corporate committee/ process; regulatory capture and knowledge
Human as AI Explainer or Interpreter	"Hallucinations"/ convincing falsehoods; Challenge of statistical blackbox; "Unjust" outcomes of algorithmic decision-making (right of review)
Human as AI Creator	Expensive to train/ retrain & run leads to market concentration; Model decay; Bias finetuning; Carbon footprint & water usage; Equitable access (including for Global South and languages)
Human as Customer-user of AI Products and Services	**GenAI use** - Synthetic data polluting internet (training data); Carbon footprint & water usage; Future of work/labour; AI arms race as part of Digital Cold War; Enter "culture wars"; Challenging nature of "reality"; AGI **GenAI "misuse"** - malicious "fake news" and other "misuse"; autonomous weapon systems

Governance of AI makers

Governance of AI use/ users

GenAI Safety risks "Censored, Bias and/or Overcorrected GenAI"?

- AI safeguards for content moderation and safety to refuse to generate certain content
  - Risk censorship/ bias / overcorrection on topics involving gender, ethnicity, hate speech, graphic violence, etc out of concern not to marginalize or offend or be misaligned with societal values via "safety classifiers" focused on toxicity, discrimination, intent and privacy
    - Eg, Allen Institute's Real Toxicity Prompts
    - Eg, Chinese TC260's "refuse to answer test question bank" (min 500 on 17 safety risks)
- "Censored GenAI" may impact utility for NGOs/ lawyers (eg, criminal law, human rights)

Twitter AI: The Role of Artificial Intelligence in Content Moderation
How does Twitter use artificial intelligence and machine learning?
TIME
SIGN UP FOR OUR ENTERTAINMENT NEWSLETTER

IDEAS • TECHNOLOGY
The Future of Censorship Is AI-Generated
6 MINUTE READ

Censorship in AI Chatbots
Gizmodo measured how AI responds to controversial questions.

The rates at which AI chatbots refused to respond to controversial questions. - Graphic Gizmodo
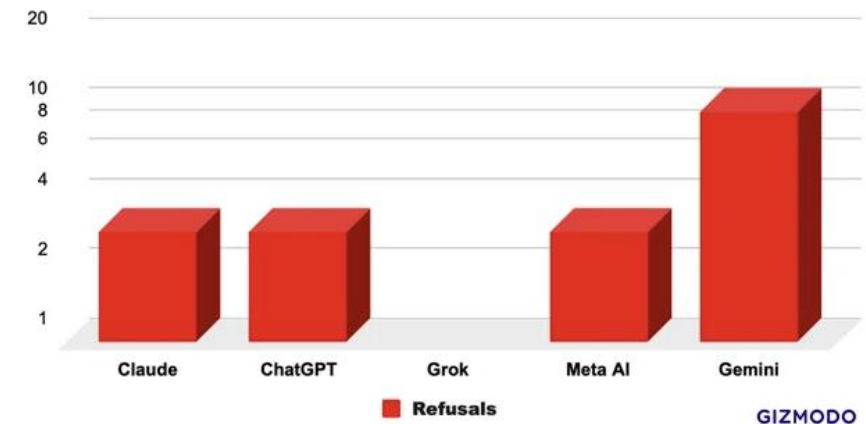
Lack of Consensus on Regulating GenAI externalities

Regulatory focus on **output risk and use of GenAI**

- High risk (including systematically risky LLMs) (China, EU; California bill)

- Employment, deep fakes, consumer protection, etc

- ISO/IEC42001 AI management system standard

Lack of consensus amongst legislators on regulating **externalities regarding how such GenAI is created**, but remain a concern for many parties

- GenAI users (corporates, governments, courts, universities, law firms & individuals)

- Employee talent of GenAI companies who are the humans-working-in-the-loop

- Investors and owners

Two Targets of Market-based Approaches

(a) Market mechanisms addressing individual consumer benefit (promised performance) and consumer protection (against individual harm)

- o "Hallucinations" / convincing falsehoods that result in individual harm
- o Statistical blackbox concerns that prevent explainability and accountability
- o "Unjust" outcomes on those prejudiced by algorithmic decision-making (including from bias training data, bias finetuning)
- o Model decay that leads to inconsistent outputs from the same inputs
- o Malicious "fake news" and other "misuse" that result in individual harm
- o Confidentiality and data protection of eg prompting and output as GenAI training

Two Targets of Market-based Mechanisms for GenAI externalities

(b) Market mechanisms addressing third-party benefit/ holistic/ ethical considerations (ie, beyond consumer protection)

- o Copyright infringement from training using copyrighted images on the internet
- o Poor data worker pay & conditions
- o Resultant market concentration due to cost to train models
- o Excessive carbon footprint, water usage and energy need from training and usage
- o Inequitable access, including from Digital Divide, Global South and languages
- o Synthetic data "polluting" internet that reduces the quality of its data and use
- o Impact on the future of work, labour and jobs
- o Exasperating an AI arms race as part of geopolitical Digital Cold War

Two Targets of Market-based Mechanisms for GenAI externalities

(b) Market mechanisms addressing third-party benefit/ holistic/ ethical considerations (ie, beyond consumer protection)

o   Exasperating "culture wars" through fine-tuning of models that reflect different perspectives (eg, gender)

o   Malicious "fake news" and other "misuse" that begins to challenge the perceived notions of "reality" in society

o   Misuse in autonomous weapon systems

o   GenAI safety measures resulting in "censored", bias and/or "overcorrected" GenAI that begins to challenge the perceived notions of "reality" in society (eg, Google Gemini "woke" images, Allen Institute's Real Toxicity Prompts, Chinese TC260's "refuse to answer test question bank" (min 500 on 17 safety risks))

o   Existential concerns about society being overrun by AGI

Market-based Approaches as "ethical consumerism" – stakeholder's choice

(1) Voluntary certification
(2) Rating and scoring schemes
(3) Nutrition label

Potentially more powerful application beyond individual/retail consumers

- Corp, govt, judicial, universities & law firm consumers of GenAI systems

- Employee talent HWITL of GenAI companies

- Investors and owners

Choose the externality that stakeholders care about

Risks of "greenwashing" or "fairwashing" from misuse of labels

(1) Voluntary certification schemes: Fair trade movement

Aim – to recognize products that assist developing country agricultural producers be fairly compensated
- Work with farmers and workers of 300 commodities since 1997

But multiple logos and standards

- "Fair Trade" USA split in 2012 due to decision to certify plantation grown coffee - "water down standards"
- Criticism for not addressing other externalities eg, environmental concerns, labor and production standards
- Criticism for not achieving equitable outcome sought

(1) Voluntary certification schemes: Fair trade movement applied to GenAI

Fairly Trained Certified

Aim – to recognize LLM products that are trained with licensed data (ie, not infringement of copyright materials)

Query - impact on performance?
Query - less relevant if copyright data mining exception ? (eg, Singapore, Japan)

KL3M, the first clean LLM.

Language models without **IP** or **toxicity** issues.

Designing your AI Solution Ethically

Brian W Tang

Founding executive director
Law, Innovation, Technology & Entrepreneurship Lab (LITE Lab)
University of Hong Kong Faculty of Law

bwtang@hku.hk