

Safe Innovation as Infrastructure: Making AI Assurance and Resilience Integral to Hong Kong’s First Five-Year Plan

*Submission by AI Safety Hong Kong to the public consultation on The First Five-Year
Plan for Economic and Social Development of the HKSAR (2026–2030)*

About AI Safety Hong Kong

AI Safety Hong Kong (AISHK) is a collective of Hong Kong residents concerned with the safe development and use of advanced artificial intelligence. We undertake study, research, public commentary and community convening activities on AI governance, assurance and resilience as they relate to Hong Kong. We can be contacted at aisafetyhongkong@gmail.com

Opening.....	2
The case: safe innovation as infrastructure.....	3
Recommendation 1 — Embed safe, secure, trustworthy and accountable AI across the Plan.....	7
How this responds to the Five-Year Plan’s own wording.....	7
Applying one principle across the Plan.....	7
Rationale.....	9
Illustrative implementation pathways.....	10
Recommendation 2 — Ensure AIRDI’s assessment and certification-related work is credible, transparent and demonstrably independent.....	11
How this responds to the Plan’s own wording.....	11
Rationale.....	11
Illustrative implementation pathways.....	12
Recommendation 3 — Build Hong Kong’s role as an AI-assurance and standards-interoperability bridge.....	13
How this responds to the Plan’s own wording.....	13
Rationale.....	13
Illustrative implementation pathways.....	15
Conclusion.....	15
Annex — Suggested Plan text and illustrative implementation pathways.....	16
A. Suggested Plan text.....	16
B. Illustrative implementation pathways.....	17

Opening

AISHK supports the Five-Year Plan's ambition to deepen the implementation of the AI+ Initiative, accelerate the transformation of AI R&D outcomes and promote AI integration across industries. We recommend that the Plan pair these ambitions with practical capacity in AI assurance, evaluation, security, resilience, professional training and standards interoperability, so that Hong Kong's AI development is trusted, internationally connected and sustainable.

Our central proposition is that Hong Kong should make AI safety, assurance and resilience integral to its AI development strategy, because safe innovation is infrastructure for trusted adoption, high-quality development, financial resilience and international connectivity.

We make three recommendations:

1. **Embed safe, secure, trustworthy and accountable AI** across the Plan's AI+, innovation, public-service and resilience commitments.
2. **Ensure that AIRDI's emerging AI assessment and certification-related work is credible, transparent and demonstrably independent** from relevant commercial advisory interests.
3. **Build Hong Kong's role as an AI-assurance and standards-interoperability bridge**, using its strengths in testing and certification, accreditation, professional services, finance, universities, Greater Bay Area cooperation and international connectivity.

The case: safe innovation as infrastructure

The Plan already contains the principle we are asking it to apply consistently. Its guiding principles commit Hong Kong to “follow a holistic approach to development and security, enhancing the resilience of the economy and society.”¹ Its discussion in Part 6 about making Hong Kong a more resilient city requires “that equal emphasis is placed on both development and safety.”² The Chief Executive’s 2025 Policy Address commits the Government to promote AI+ and deep integration of AI across sectors “while placing strong emphasis on safety risk prevention.”³ In June 2026 the Innovation, Technology and Industry Bureau told the Legislative Council that “AI development must be guided by safety and driven by applications.”⁴

Our present submission asks that the Plan’s AI commitments be drafted to this same standard. At present, the AI+ commitment in Part 3 is expressed as AI diffusion — accelerating the transformation of R&D outcomes, developing application scenarios, and promoting deep integration across industries. Its only explicit governance element is a single general clause: “refine the framework for AI and data governance.”⁵ The clause does not identify any of the specific capabilities — such as evaluation, testing or assurance — on which a refined framework would rest. This matters because a five-year plan shapes what follows it: capabilities the Plan names are more likely to be given an owner, a budget and performance indicators; capabilities it leaves unnamed may not be.

Assurance can support adoption. AI assurance is the practical work of establishing, through testing, audit and documented evidence, that an AI system or the organisation deploying it meets defined requirements. Organisations deploying AI in regulated or high-impact settings need to satisfy boards, auditors, insurers and supervisors that a

¹Constitutional and Mainland Affairs Bureau, The First Five-Year Plan for Economic and Social Development of the Hong Kong Special Administrative Region (2026–2030): Public Consultation Document (2026), Background, p.2. Cited below as “Public Consultation Document”. <https://www.hk5yplan.gov.hk/>

²Public Consultation Document, Part 6, p.32.

³The Chief Executive’s 2025 Policy Address, para. 66. <https://www.policyaddress.gov.hk/2025/en/p66.html>

⁴Innovation, Technology and Industry Bureau and Digital Policy Office, “Work on Promoting Artificial Intelligence Ecosystem Development and the ‘AI for All’”, paper for the Legislative Council Panel on Information Technology and Broadcasting, LC Paper No. CB(2)730/2026(02), 8 June 2026, para. 19. Cited below as “LC Paper No. CB(2)730/2026(02)”. <https://www.legco.gov.hk/yr2026/english/panels/itb/papers/itb20260608cb2-730-2-e.pdf>

⁵Public Consultation Document, Part 3, p.14.

system is controlled. Shared assurance tools and methods can make that easier to evidence than approaches developed firm by firm.

Hong Kong’s financial regulators already use risk-controlled testing to support responsible AI innovation: in March 2026, the Monetary Authority, the Securities and Futures Commission (SFC), the Insurance Authority and the Mandatory Provident Fund Schemes Authority, together with Cyberport, expanded the GenA.I. Sandbox launched with the banking sector in 2024 into the GenA.I. Sandbox++, spanning banking, securities and capital markets, asset and wealth management, insurance, MPF and stored-value facilities, with a focus on risk management, anti-fraud and customer experience.⁶

Financial resilience. The Plan seeks to strengthen Hong Kong’s functions as an “international risk management centre” and to deepen the development of insurance and risk management markets.⁷ As AI is adopted across financial services, the risks that accompany it — model risk, reliance on third-party providers, AI-enabled fraud, and the operational resilience of AI-dependent services — are relevant to that role, and assurance capacity supports it. Supervisory expectations already reflect this. The SFC requires licensed corporations to maintain effective AI model risk management, to validate models before approving them for use and when material changes are made, to assess performance through comprehensive end-to-end testing covering the entire process from user input to system output, to keep model performance under ongoing review and monitoring, and to address cybersecurity, data and third-party provider risk.⁸

Assurance builds on strengths Hong Kong already has. The Plan names “testing and certification” among the supporting services whose synergised development the

⁶Hong Kong Monetary Authority, Securities and Futures Commission, Insurance Authority, Mandatory Provident Fund Schemes Authority and Cyberport, “Regulators launch GenA.I. Sandbox++ to foster A.I. innovation across financial services”, press release, 5 March 2026.

<https://www.hkma.gov.hk/eng/news-and-media/press-releases/2026/03/20260305-3/>

⁷Public Consultation Document, Part 2, pp.9–10.

⁸Securities and Futures Commission, “Circular to licensed corporations — Use of generative AI language models”, 12 November 2024 (ref. 24EC55), paras. on AI model risk management, cybersecurity and data risk management, and Third Party Provider risk management.

<https://apps.sfc.hk/edistributionWeb/api/circular/list-content/circular/intermediaries/supervision/doc?refNo=24EC55&lang=EN>

Plan seeks, alongside trade finance, insurance, legal services and arbitration.⁹ The Hong Kong Accreditation Service (HKAS), under the Innovation and Technology Commission, operates accreditation schemes for laboratories, certification bodies and inspection bodies, and participates in international mutual-recognition arrangements.¹⁰ Hong Kong could use HKAS's existing accreditation system to support the accreditation of bodies that test, inspect, or certify AI systems for safety. This would build on an established institutional function, while still requiring new AI-assurance standards and assessment criteria.

Resilience serves a distinct purpose. As AI is integrated more deeply into industries, public services and infrastructure, the city's risk assessment, early warning and contingency arrangements can usefully take account of it — including cyber misuse, AI-enabled fraud, the failure of increasingly autonomous systems, and the dependence of essential services on AI providers. The Five-Year Plan commits to “comprehensively strengthen security governance of the city, including public safety and cybersecurity,” and to inter-departmental collaboration, risk assessment, early warning mechanisms and contingency plans.¹¹ Adding AI explicitly to that work would apply the Plan's own principle — equal emphasis on development and safety — to the technology the Plan most seeks to promote.

International connectivity. The Plan commits Hong Kong to advance “soft connectivity” in rules and standards,¹² to enhance its participation in “professional standards ... and global governance,”¹³ and to “serve as a bridge connecting with the Chinese Mainland and the world, and facilitate Chinese rules and standards to ‘go global’.”¹⁴ Assurance and standards work is a concrete, technical way of doing this. It

⁹Public Consultation Document, Part 2, p.11.

¹⁰Innovation and Technology Commission, “Hong Kong Accreditation Service — About HKAS”.
<https://www.itc.gov.hk/en/quality/hkas/about/hkas.html>

¹¹Public Consultation Document, Part 6, p.32.

¹²Public Consultation Document, Part 5, p.24.

¹³Public Consultation Document, Part 5, p.25.

¹⁴Public Consultation Document, Part 5, p.26.

aligns with Hong Kong's established roles in testing and certification, financial and risk-management services, and professional services.

Recommendation 1 — Embed safe, secure, trustworthy and accountable AI across the Plan

We recommend that the Five-Year Plan provide that the deep integration of AI across industries and public services is to proceed on a safe, secure, trustworthy and accountable basis, and that its commitment to refine the framework for AI and data governance be expanded to name the capabilities intended to deliver it — including AI evaluation, testing, assurance, security and incident learning.

How this responds to the Five-Year Plan’s own wording

The Plan commits to “[d]eepen the implementation of the AI+ Initiative, accelerate the transformation of AI R&D outcomes and development of application scenarios, promote the deep integration of AI across various industries, and refine the framework for AI and data governance.”¹⁵ Our recommendation does not change the direction of that commitment. It asks the Plan to say, in its own text, what the Government has already said elsewhere: the 2025 Policy Address commits to promoting AI “while placing strong emphasis on safety risk prevention”,¹⁶ and the Innovation, Technology and Industry Bureau told the Legislative Council in June 2026 that “AI development must be guided by safety and driven by applications”.¹⁷ It also gives the governance clause enough specificity to be acted upon. Suggested drafting to incorporate our recommendation within the Plan is set out in the **Annex**.

Applying one principle across the Plan

The table below shows how the single cross-cutting principle in Recommendation 1 applies to the AI-facing commitments already present in each Part of the Five-Year Plan. It is one recommendation applied across the whole Plan.

¹⁵Public Consultation Document, Part 3, p.14.

¹⁶The Chief Executive’s 2025 Policy Address, para. 66.

¹⁷LC Paper No. CB(2)730/2026(02), para. 19.

Plan area	How the safe-AI principle applies
Part 1 — Northern Metropolis, University Town, Hetao	The I&T industry chain is to encompass “platforms for R&D, testing, manufacturing, demonstration and commercialisation.” ¹⁸ AI assurance and evaluation capability belongs within that chain, and it also belongs among the multi-disciplinary areas developed in the University Town.
Part 2 — International financial centre and professional services	Strengthening Hong Kong as an “international risk management centre” ¹⁹ can encompass the risks that accompany AI adoption in financial services; “testing and certification” ²⁰ can extend to AI.
Part 3 — AI+, commercialisation and AI/data governance	The commitments to deepen the AI+ Initiative, accelerate the transformation of R&D outcomes and promote deep integration across industries, ²¹ together with the optimisation of testing platforms and laboratory facilities, should proceed on a safe, secure, trustworthy and accountable basis.
Part 4 — Education, skills, labour and livelihood	Smart healthcare and healthcare information infrastructure ²² may involve high-impact AI use. Training and upskilling commitments ²³ can include content on AI safety, security, governance and assurance content.

¹⁸Public Consultation Document, Part 1, p.7.

¹⁹Public Consultation Document, Part 2, pp.9–10.

²⁰Public Consultation Document, Part 2, p.11.

²¹Public Consultation Document, Part 3, p.14.

²²Public Consultation Document, Part 4, p.19.

²³Public Consultation Document, Part 3, p.15; Part 4, p.20.

Part 5 — GBA, rules-and-standards connectivity, Belt and Road	Rules-and-standards connectivity ²⁴ can include AI standards and assurance, as developed below in Recommendation 3.
Part 6 — City resilience, public safety and cybersecurity	Security governance, inter-departmental risk assessment, early warning and contingency planning ²⁵ can take account of AI risks and related considerations, as can the development of autonomous vehicles and smart city infrastructure. ²⁶

The Five-Year Plan’s resilience commitments in Part 6 are drafted principally around natural hazards and emergencies. As AI is integrated more deeply into infrastructure and essential services, we suggest that risk assessment, early warning and contingency arrangements take account of AI-related considerations — including cyber misuse, AI-enabled fraud, the failure of increasingly autonomous systems, and dependence on AI services — and that meaningful human accountability be maintained for consequential decisions in public services.

Rationale

Hong Kong has capable governance instruments. The Digital Policy Office maintains the Ethical Artificial Intelligence Framework²⁷ and the Hong Kong Generative Artificial Intelligence Technical and Application Guideline,²⁸ which address matters including system security and robustness, personal privacy, and data security and management, and the Government reports that it regularly examines AI risks, safe application and governance standards.²⁹ The Secretary for Justice has convened a steering committee to

²⁴Public Consultation Document, Part 5, p.24.

²⁵Public Consultation Document, Part 6, p.32.

²⁶Public Consultation Document, Part 6, pp.30–31.

²⁷ DPO, “Ethical Artificial Intelligence Framework”, https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/

²⁸ DPO, “Hong Kong Generative Artificial Intelligence Technical and Application Guideline”, https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/doc/HK_Generative_AI_Technical_and_Application_Guideline_en.pdf

²⁹LC Paper No. CB(2)730/2026(02), paras. 20–21.

establish an inter-departmental working group that will review legislation to support the wider application of AI.³⁰ Public-sector deployment is proceeding at pace under the AI Efficacy Enhancement Team, which aims to introduce AI tools to 100 administrative procedures within 2026 and at least 200 by end-2027.³¹

What we have not identified is a Plan-level commitment to the underlying capacity — the ability to evaluate, test and assure AI systems, to develop the professional skills this requires, and to learn systematically from incidents. Guidance describes good practice. Capacity is what allows good practice to be checked, evidenced and improved.

This matters most where the Government is itself the deployer. As AI tools are introduced into administrative procedures, proportionate assurance of the Government's own AI procurement and high-impact public-sector use would support public confidence and provide a practical model for other jurisdictions.

Illustrative implementation pathways

- A published account of how AI assurance, evaluation and incident-learning capacity will be developed over the Plan period.
- A cross-sector design study and pilot of voluntary, confidential AI incident-learning arrangements, coordinated with existing privacy, cybersecurity, critical-infrastructure and sectoral reporting channels.
- Review of existing AI guidance and/or the creation of new binding guidelines to address increasingly autonomous systems, including permission boundaries, logging and points of human intervention.
- Proportionate assurance arrangements for the Government's own AI procurement and high-impact public-sector deployments.

³⁰LC Paper No. CB(2)730/2026(02), para. 22.

³¹LC Paper No. CB(2)730/2026(02), para. 14.

Recommendation 2 — Ensure AIRDI’s assessment and certification-related work is credible, transparent and demonstrably independent

We recommend that, as AIRDI’s operating arrangements are settled, its AI assessment and certification-related work be supported by arrangements that make its independence demonstrable — including appropriate separation of assessment work from fee-earning advisory activity, transparent assessment methodologies, published conflicts-of-interest management, and safety- and assurance-related performance measures.

How this responds to the Plan’s own wording

The Five-Year Plan seeks to refine the framework for AI and data governance³² and to develop testing and certification as a supporting professional service.³³ AIRDI could play a central role to both goals. AIRDI’s stated objectives include establishing AI standards and promoting system interoperability, providing AI safety assessment and related consulting services, and building cross-sector and inter-regional collaboration platforms to promote the standardisation and safe application of AI.³⁴

Rationale

We support AIRDI’s public mission, and we note the substantial accountability arrangements that already apply to it as a wholly government-owned company — a government-appointed Board, a designated Controlling Officer, unimpeded audit access, and regular reporting to the Legislative Council.³⁵ Our recommendation concerns one specific design question.

³²Public Consultation Document, Part 3, p.14.

³³Public Consultation Document, Part 2, p.11.

³⁴“LCQ6: Regulation and development of AI technology”, Legislative Council, 18 March 2026.
<https://www.info.gov.hk/gia/general/202603/18/P2026031800419.htm>

³⁵Finance Committee paper FCR(2025-26)46, “Hong Kong Artificial Intelligence Research and Development Institute”, Legislative Council Finance Committee, September 2025, paras. 13, 16 and 27–28, pp.5, 8–9 and 12. Cited below as “FCR(2025-26)46”. <https://www.legco.gov.hk/yr2025/english/fc/fc/papers/f25-46e.pdf>

The published funding proposal contemplates that AIRDI may develop rating and certification-related business providing AI system security and ethical assessment, together with compliance risk assessment and certification, and may establish an AI experimental testing platform.³⁶ That work is described within the same business direction as technical advisory and consulting services provided on a fee-charging basis, and the proposal contemplates diversified funding including service fees.³⁷

The Government appointed AIRDI's Board of Directors in March 2026, and AIRDI is expected to commence full operation in the second half of 2026.³⁸ Its assessment work is therefore still being designed. The value of an assessment to a third party user depends substantially on how visibly independent that assessment is of the commercial interests connected to it. Settling arrangements that make independence demonstrable at the design stage is simpler than adjusting them later, and would strengthen the credibility of assessments issued in Hong Kong for local, Mainland and international users alike.

We recognise that AIRDI's final operating arrangements may not yet be public and may already address some of these considerations; these recommendations are offered to strengthen the credibility of its assurance functions as they are settled.

Illustrative implementation pathways

- Appropriate separation between assessment activity and fee-earning advisory services, reflected in AIRDI's operating and governance arrangements.
- Publication of assessment methodologies and of arrangements for managing conflicts of interest.
- Inclusion of safety- and assurance-related measures among the performance measures against which AIRDI's work is reported.

³⁶FCR(2025-26)46, para. 15(c)(ii), pp.7–8.

³⁷FCR(2025-26)46, para. 15 chapeau, p.6; para. 15(c)(iii), p.8.

³⁸“Government announces appointments to Board of Directors of Hong Kong Artificial Intelligence Research and Development Institute”, press release, 12 March 2026, <https://www.info.gov.hk/gia/general/202603/12/P2026031200378.htm>; LC Paper No. CB(2)730/2026(02), para. 7.

Recommendation 3 — Build Hong Kong’s role as an AI-assurance and standards-interoperability bridge

We recommend that the Five-Year Plan develop Hong Kong’s role as a bridge for AI assurance and standards interoperability, drawing on its testing and certification sector, accreditation infrastructure, professional services, financial sector, universities, Greater Bay Area cooperation and international connectivity — with cross-boundary cooperation and recognition arrangements explored subject to technical, legal and institutional feasibility.

How this responds to the Plan’s own wording

The Five-Year Plan commits to advance “soft connectivity” in rules and standards and to deepen the alignment of rules and mechanisms in the Greater Bay Area;³⁹ to develop the Hetao Hong Kong Park as a platform for high-level R&D collaboration and outcome transformation, and to bring together “capital, projects, talents, technology, standards and professional services” for Belt and Road cooperation;⁴⁰ and to serve as a two-way bridge helping Chinese rules and standards “go global.”⁴¹ It also commits to develop in the Northern Metropolis “a more comprehensive I&T industry chain, encompassing platforms for R&D, testing, manufacturing, demonstration and commercialisation,”⁴² and to strengthen applied education and skills training.⁴³

Rationale

Organisations deploying AI across Hong Kong, the Mainland and international markets encounter differently structured requirements. Work that maps these requirement sets to one another, and that develops the professional competence to apply them, is practical, technical and well-matched to Hong Kong’s institutions. It is also directly responsive to the Plan’s own connectivity commitments.

³⁹Public Consultation Document, Part 5, p.24.

⁴⁰Public Consultation Document, Part 5, p.25.

⁴¹Public Consultation Document, Part 5, p.26.

⁴²Public Consultation Document, Part 1, p.7.

⁴³Public Consultation Document, Part 3, p.15; Part 4, p.20.

Hong Kong's accreditation and testing infrastructure is the natural foundation for this role. Accreditation can provide independent evidence of a certification body's competence and impartiality and may strengthen the credibility and potential international usability of its certificates, subject to the relevant accreditation arrangements, market acceptance and applicable legal requirements. Whether, and on what basis, AI conformity assessment results issued in Hong Kong could be recognised elsewhere is a question worth exploring.

Capacity depends on people. Assurance work requires assessors, auditors, model-risk practitioners and AI-literate compliance and legal professionals. The Government's AI for All Inclusive Programme provides an emerging vehicle for public, professional and workplace AI training: the June 2026 LegCo paper describes tiered activities led by Cyberport, HKSTP and HKPC, including professional and workplace-focused training.⁴⁴ Incorporating safety, security, governance and assurance content into these programmes would begin building the assessor, auditor and model-risk skills base upon which any assurance activity depends. The Digital Policy Office, with the Hong Kong Internet Registration Corporation, is also launching a Safe AI@Workplace Empowerment Programme offering practical material including assessment checklists.⁴⁵

A distinction worth stating. Conformity assessment and advanced-model capability evaluation are distinct but complementary. Conformity assessment establishes whether an organisation, system or process meets defined governance, documentation, testing, security or management requirements; it is auditable and can be accredited.

Independent capability evaluation examines what an advanced model can do, including in relation to cyber misuse, dangerous scientific misuse, autonomous tool use and the maintenance of meaningful human oversight; methods for it are still developing internationally. A conformity-assessment certificate does not necessarily establish that a model lacks particular capabilities. Hong Kong is well placed to develop the first during this Plan period, and to build research capacity in the second through its universities.

⁴⁴LC Paper No. CB(2)730/2026(02), paras. 23–24.

⁴⁵LC Paper No. CB(2)730/2026(02), para. 33.

Illustrative implementation pathways

- Technical mapping among applicable national and international AI standards and requirements, maintained as a public resource.
- Exploration of accredited AI conformity assessment, and of recognition pathways, subject to technical, legal and institutional feasibility.
- Inclusion of safety, security, governance and assurance content within publicly supported AI training, and support for university research and evaluator training. This also extends to AI safety awareness training for civil servants and other officials, in addition to basic AI literacy training.

Conclusion

Hong Kong's AI ambitions are substantial and encouraging. The Plan already states the principle that development and safety deserve equal emphasis,⁴⁶ and the Government has said that AI development must be guided by safety and driven by applications.⁴⁷ We ask that this principle be carried into the Five-Year Plan's AI commitments themselves — by naming the assurance, evaluation, security and resilience capacity that gives it effect, by settling AIRDI's assessment arrangements in a way that makes their independence demonstrable, and by developing Hong Kong's role as a bridge for AI assurance and standards interoperability.

We would be glad to provide further detail on any of these recommendations.

⁴⁶Public Consultation Document, Part 6, p.32.

⁴⁷LC Paper No. CB(2)730/2026(02), para. 19.

Annex — Suggested Plan text and illustrative implementation pathways

A. Suggested Plan text

Part 3 — AI+, commercialisation and AI/data governance

Current wording:⁴⁸

“Deepen the implementation of the AI+ Initiative, accelerate the transformation of AI R&D outcomes and development of application scenarios, promote the deep integration of AI across various industries, and refine the framework for AI and data governance.”

Suggested wording:

*“Deepen the implementation of the AI+ Initiative, accelerate the transformation of AI R&D outcomes and development of application scenarios, and promote the deep integration of AI across various industries **and public services on a safe, secure, trustworthy and accountable basis**; refine the framework for AI and data governance, **including by strengthening the capabilities on which effective governance depends — among them AI evaluation, testing, assurance, security and incident learning.**”*

Part 6 — City resilience, public safety and cybersecurity (optional addition)⁴⁹

“In strengthening city resilience, public safety and cybersecurity, take account of risks and dependencies arising from the adoption of AI in essential services, public services and critical systems, and strengthen relevant risk assessment, early warning, contingency and human-accountability arrangements.”

⁴⁸Public Consultation Document, Part 3, p.14.

⁴⁹Public Consultation Document, Part 6, p.32.

B. Illustrative implementation pathways

These pathways illustrate how the recommendations could be given effect. They are indicative rather than prescriptive.

Standards mapping and conformity-assessment pathways. A maintained public mapping among applicable national and international AI standards & requirements would identify where they coincide, where they diverge, and what additional evidence each calls for. Such a resource would support organisations operating across markets and would be a concrete contribution to rules-and-standards connectivity.

Building on this, Hong Kong could assess whether, and on what basis, accredited AI conformity assessment might be offered here, drawing on the existing accreditation infrastructure and applicable international requirements for certification bodies. Accreditation can provide independent evidence of a certification body's competence and impartiality, but it does not by itself guarantee that certificates will be accepted across borders; recognition depends on the relevant accreditation arrangements, market acceptance and applicable legal requirements, and would need to be explored subject to technical, legal and institutional feasibility.

Greater Bay Area and Hetao pilots. Pilot applications of assurance approaches in high-impact areas, hosted through existing cooperation platforms (such as the GBA Joint Laboratory for Generative AI Security Development),⁵⁰ would produce documented and replicable methods. Pilots offer a low-risk way of testing whether the standards mapping works in practice, and of identifying where divergences create friction for organisations operating on both sides of the boundary.

Training, research and technical capacity. Safety, security, governance and assurance content could be incorporated within publicly supported AI training. Practitioner and assessor curricula could be developed with the vocational sector, professional bodies and universities, building the competence base that any assurance activity depends on. University research and fellowships on evaluation methodology, agent security and model-risk methods would develop the deeper technical capacity that conformity assessment alone does not provide, and would give Hong Kong a basis for contributing to international technical work in this field.

⁵⁰ Guangzhou Municipal People's Government, '大湾区生成式人工智能安全发展联合实验室9月11日成立' ['Greater Bay Area Generative Artificial Intelligence Safety and Development Joint Laboratory Established on 11 September']. https://www.gz.gov.cn/zwfw/zxfw/kjcy/content/post_9865069.html

AIRDI assurance safeguards. Separation of assessment work from fee-earning advisory activity; transparent assessment methodologies; published arrangements for managing conflicts of interest; and safety- and assurance-related measures among the performance measures against which AIRDI's work is reported. Adopting such arrangements while AIRDI's assessment work is still being designed would be simpler than introducing them once assessment services are established.