Which AI risks are Hong Kong's financial regulators aware of?

Summary

We propose to map what Hong Kong's financial regulators have said about AI and evaluate their engagement with catastrophic AI risks, by evaluating their policy documents against the risk categories of the MIT AI Risk Repository. The regulators have been writing about AI since at least 2019, and the rate of output has risen in recent years, especially in 2026. The Hong Kong Monetary Authority replaced its AI principles with a new supervisory framework in October, the HKMA and the Securities and Futures Commission issued circulars on AI-enabled cyberattacks in June, the four financial regulators ran a joint sandbox for AI agents, and the HKMA published a discussion paper on agentic AI in November. Nobody so far has evaluated these documents together against a full list of AI risks.

The team will first decide which of MIT's 24 categories fall within a financial regulator's remit, then compile every document on AI issued by the regulators and read each one against those categories, recording for every match the passage concerned, how binding the document is, and whether it refers to AI directly or through a broader concept such as model risk. We will publish a report, a dataset and a one-page map of regulators against risks. The project will start with Hong Kong, so that we have a minimum viable output by the end of the programme. If the team is efficient enough, we will add Singapore as a comparison. This is the first stage of a broader research project on AI safety across the major financial centres.

Background

The MIT AI Risk Repository is a database of more than 1,700 risks drawn from over 65 published frameworks and organised into 24 categories, which MIT calls subdomains. As far as we know it is the most comprehensive list of AI risks available. MIT also uses the taxonomy to classify laws and policies. Its AI Governance map covers more than 1,000 documents from CSET's AGORA archive, classified by language models. However, the content is mostly US-centric, and as far as we can tell no Hong Kong financial regulator appears in the map or in the archive behind it. In the world of financial regulation, the OECD and the Bank for International Settlements have surveyed AI rules across many jurisdictions, and law firms summarise each new Hong Kong circular, but all of this describes documents one at a time. We have not found any study that reads the full set of Hong Kong documents against a comprehensive list of AI risks.

Hong Kong is worth studying on its own. It is Asia's top financial capital and the world's top cross-border wealth hub, so what its regulators require of AI use in finance applies to a large share of global financial activity. It is the listing venue for Chinese frontier AI developers doing global IPOs, and it is where Mainland China's AI firms meet non-Chinese regulatory expectations and international capital.

Our central question for the project: for each MIT risk category that a financial regulator could address, which Hong Kong financial regulators have demonstrable awareness of it, in which document, how binding is that document, and what does it say?

Theory of change

Financial regulators have no say in how AI systems are built. However, they can decide what the firms that use those systems must do, and in many jurisdictions they are the only bodies with legal power over the use of AI today. If a financial regulator has not recognised a risk in writing, no firm it supervises will be asked to manage that risk.

The MIT taxonomy includes several categories that the AI safety field regards as potentially catastrophic: AI with dangerous capabilities, AI pursuing goals that conflict with human goals, risks from many AI agents interacting, cyberattacks and weapons, competitive pressure to deploy unsafe systems, and concentration of power. These tend to sit outside the scope of what financial regulators typically enforce, which is why it is worth asking whether Hong Kong's regulators have engaged with them at all. The map will show, for each regulator, which risks it has addressed, which its peers have addressed, and which none of them have.

This project is the first building block of a broader research programme. The same method can be applied to the other major financial centres, which would allow us to test the argument in Sijuade Animashaun's work on cooperation between global financial centres, most recently in "Crypto Badlands", that the leading centres should coordinate their regulation of new technology. The findings will also have practical value for the Hong Kong AI Safety Hub's planned convening activities with Hong Kong's financial regulators.

Research plan

Hong Kong follows a sectoral model of financial regulation. Four statutory regulators each supervise one part of the sector under their own ordinance:

- the HKMA for banks under the Banking Ordinance,
- the SFC for securities and futures under the Securities and Futures Ordinance,
- the Insurance Authority under the Insurance Ordinance, and
- the Mandatory Provident Fund Schemes Authority under the Mandatory Provident Fund Schemes Ordinance.

They are coordinated by the Council of Financial Regulators, chaired by the Financial Secretary, alongside the Financial Services and the Treasury Bureau, which sets financial policy for the government. The study will cover all of these bodies, together with HKEX, which the Securities and Futures Ordinance makes the frontline regulator of listed companies. It will also cover three bodies outside the financial sector whose rules bind financial firms: the Privacy Commissioner for Personal Data, the Digital Policy Office, and the regime under the Protection of Critical Infrastructures (Computer Systems) Ordinance. Without these, a risk that no financial regulator has addressed might wrongly appear to be unaddressed in Hong Kong altogether.

In the first two weeks the team will decide, with the co-leads, which of the 24 MIT categories a financial regulator could address under its statutory functions, and will record the reasons behind this selection. Categories that fall outside every regulator's remit will be noted in the report but will not be mapped.

The team will then compile every document that is currently in force on AI from the bodies covered, including ordinances, rules, codes, guidelines, circulars, and policy statements, and will archive a copy of each.

Each document will then be read against the categories shortlisted from the MIT AI Risk Initiative. For every match, the researchers will record the passage and a supporting quotation, whether the document refers to AI directly or through a broader concept, how binding the document is (a rule, guidance, or a statement), its status, and a link to the archived copy. A passage may be coded under more than one category. The team will review one another's coding at the weekly meeting and will agree the final coding of every document by consensus, with the co-leads advising on questions of law and on the meaning of the MIT categories. For each category, the report will also note the corresponding vulnerability identified by the Financial Stability Board in its 2024 report on AI and financial stability, so that regulators can relate the findings to the framework they already use.

Phase	Weeks	Work
Set-up	1 to 2	Read the MIT paper and Crypto Badlands; decide which categories a financial regulator could address; agree the coding sheet; code two documents together as practice.
Compile	3 to 5	Compile and archive the documents for each body.
Code	6 to 10	Code every document; review coding at the weekly meeting.
Review	11 to 12	Agree the final coding of every document; build the map.

Write	12 to 14	Draft the report and the presentation; publish the dataset.

Team

We will need at least two researchers. If the work goes faster than expected, or we recruit three strong researchers, we may add Singapore to give the project an Asian comparative frame.

Researchers will each take a group of bodies, compile and archive their documents, code them, and take part in the weekly review. The work suits people who read regulatory or legal text carefully and handle sources well. A background in law, financial regulation or public policy is helpful, and we would like at least one team member to have it. No advanced background in AI safety is needed, beyond some minimum context that any suitable intro to AI governance course could provide. Team members are expected to spend about 10 hours a week on the project.

Sijuade Animashaun and Abeer Sharma are the co-leads. They will run the weekly meetings, supervise the research, and edit the outputs.

Sijuade Animashaun is a Global Research Assistant Professor at the School of Law, City University of Hong Kong. He holds a PhD in FinTech and International Financial Regulation from the University of Hong Kong and an LLM with Distinction from the University of Glasgow, which he attended as a British Chevening Scholar. His research covers international financial regulation, digital assets and coordination among financial centres, and he is the author of Crypto Badlands. He will supervise the financial regulation side of the project and will spend at least 5 hours a week on it.

Abeer Sharma is a lawyer with multijurisdictional experience and holds a PhD in Law from the University of Hong Kong. He has advised sovereign states, regulators and technology companies, works on AI governance, institutional design, and international and constitutional law, and is a co-founder of the Hong Kong AI Safety Hub. He will supervise the AI safety side of the project and will spend at least 8 hours a week on it.

Output

A report of 2,500 to 3,500 words, and a dataset with the one-page map of regulators against risk categories. These will inform future research on the other financial centres and convenings with regulators. We will also submit the findings to a journal in law or financial regulation. All team members will be co-authors, with contributions specifically highlighted in the report.

Risks and limitations

The main risk is that the map functions as little more than a scorecard. A regulator that has written something about a risk has not necessarily acknowledged it or done anything about it, and a risk category with several circulars dealing with it may reflect compliance paperwork more than attention to the risk. We will not publish scores, and every output will keep the distinction between what the regulators have said and whether it is sufficient.

A second risk is that we misread a passage and mislead a regulator or a researcher, which is why every entry will be sourced with a quotation and the team will converge unanimously on the interpretation of each document.

The map will be limited to public documents in English issued up to 28 February 2027. The research will record regulatory awareness expressed in public documents and nothing more. Whether a document shows awareness of a category is a judgment, and we will publish the reasons for each one.