Hong Kong Applied Science and Technology Research Institute
香港應用科技研究院

Supported by
保險業監管局
Insurance Authority

Whitepaper on

Federated Learning: Unlocking Innovation in the Insurance Sector

November 2025

Acknowledgements

An Advisory Panel of experts, industry players, and business stakeholders provided advice on the development of the advanced federated learning (FL) application for insurers in Hong Kong. Its members have made significant contributions to the applied research presented in this white paper, actively participating in working group discussions, sharing their knowledge and experience, and collaborating in experiments and Proof-of-Concept (PoC) implementations. We sincerely thank the following members for their valuable insights and their feedback on the white paper, while also expressing our appreciation to other members for the time they devoted to attending the panel meetings.

Member of the Advisory Panel (In alphabetical order)	Representative	Title
BOC Life Assurance Company Limited	Mr. Alfred Cheung	Deputy Chief Executive & Chief Sales Officer
Bowtie Life Insurance Company Limited	Mr. Michael Chan	Co-founder & CEO
China Taiping Life Insurance (Hong Kong) Company Limited	Mr. Jacob Wong	Executive Advisor
Finnovise	Mr. Dick Fong	CEO and Co-founder
Gilt Chambers	Mr. Stephen Wong Kai-yi	Barrister-at-law
Standard Chartered Bank	Mr. Andrew Sim	Executive Director, Head of Product and Segment Management
Swiss Re Asia Pte. Ltd., Hong Kong Branch	Mr. Jacky Ng	Data Science Director APAC
Taylor's University	Dr. Shum Kam Hong	Adjunct Professor
The Chinese University of Hong Kong	Prof. Chan Chun Kwong	Professor of Practice in Financial Technology
The Hong Kong Federation of Insurers	Ms. Selina Lau	Chief Executive
Zurich Insurance (Hong Kong)	Mr. Eric Hui	Chief Executive Officer

Contents

Contents

Executive Summary

Introduction

Data plays a crucial role in the insurance sector and in digital transformation. However, data privacy concerns and strict data handling regulations are hindering the development of data-driven solutions and AI innovations in the industry. Federated learning (FL) offers a promising solution to this problem, as it allows insurers to leverage machine learning (ML) while safeguarding individual data privacy.

Against this background, in March 2023 the Insurance Authority (IA) and the Hong Kong Applied Science and Technology Research Institute (ASTRI) undertook this research project with the aim of exploring some potential FL applications for the insurance industry.

The project had three stages:

- **Stage 1. Platform Development:** ASTRI developed an FL platform specifically tailored for the insurance industry.
- **Stage 2. Proof-of-Concept (PoC):** This stage involved data collaboration between insurers and various other sectors in order to evaluate the platform's efficiency and effectiveness.
- **Stage 3. White Paper:** This white paper documents the PoC stage and its findings, and discusses various FL-related technical risks and compliance issues.

Objectives

This white paper aims to:

- Enhance the insurance industry's understanding of FL, with a focus on its potential to help the industry leverage alternative data effectively;
- Identify and address technical risks, compliance considerations, and governance issues related to the implementation of FL within the insurance sector; and
- Describe some PoC applications that utilise FL to extract insights from diverse data sources across the insurance value chain.

Federated Learning

FL is an advanced ML technique that enables models to be trained on decentralised datasets. Unlike traditional ML which requires data to be centralised in a single location, FL enables models to be trained directly on the devices or servers where the data resides. This decentralised approach prioritises data privacy protection, enabling compliance with data privacy regulations while facilitating data collaboration across organisations and sectors. For industries that handle a lot of sensitive customer data, such as the insurance sector, FL could be a useful tool in improving operational efficiency and protecting data privacy when processing data.

Rather than adopting an open-source framework, this research project has involved developing an FL platform specifically tailored to the needs of the insurance sector. It incorporates advanced privacy-enhancing techniques, optimised algorithms, and robust modular architectures for improved security, efficiency, and flexibility. Evaluation results have demonstrated the platform's effectiveness in improving both levels of data protection and model performance.

As FL technology continues to advance, critical challenges remain that include handling diverse data types, ensuring efficient data preparation and processing, and facilitating seamless communication across systems. These issues are further examined in the PoC section.

Risks Management and Regulatory Compliance

Significant amounts of personal customer information are collected and processed by the insurance sector, making effective risk management and regulatory compliance in data handling crucial.

This paper identifies three primary types of risk commonly encountered in FL: data privacy risks, model security risks, and performance risks. Solutions for mitigating data privacy risks include secure data storage, robust authentication, and data minimisation techniques. Model security risks, or vulnerabilities to adversarial attacks, necessitate the use of defences such as differential privacy (DP) and anomaly detection. Performance risks, which stem from data heterogeneity and communication inefficiencies, can be addressed by implementing data preprocessing and optimisation strategies.

The paper also considers major compliance issues relevant to the use of FL in the insurance industry, including compliance in areas such as data privacy and protection, cybersecurity, governance frameworks, outsourcing risks, and fair customer treatment.

Finally, this paper also addresses ethical issues relevant to the responsible use of FL, including accountability and responsibility, human oversight, transparency and interoperability, fairness, robustness, safety, and security.

This paper proposes a framework for risk management, compliance, and ethical standards that is designed to serve as a starting point for stakeholders. It should enhance stakeholders' understanding of best FL practices and ensure responsible data usage, while fostering greater trust in FL applications within the insurance industry.

Proof-of-Concept

Three practical use cases have been completed in the PoC, involving three insurers and three companies from different sectors. The first use case leveraged engagement data to enhance the accuracy of an AI model for identifying potential customers. The second case incorporated clinical data to forecast the probability of insurance claims. The third utilised credit data to forecast customer renewal probabilities.

The PoC results highlighted several benefits of FL for the insurance industry. First, FL enables smarter predictive models to be developed by integrating diverse data sources, without compromising data privacy. This integration can lead to improved accuracy in predicting claims and customer behaviour, in turn supporting better pricing, resource allocation, and marketing strategies. Second, FL facilitates secure, cross-sector collaboration by allowing institutions to jointly train models without sharing sensitive data. This helps overcome data silos and regulatory barriers, and delivers richer insights and more robust models. Finally, by keeping data decentralised, FL aligns with evolving regulatory standards and fosters responsible AI practices, enhancing data privacy and reinforcing customer trust in the use of their data. Overall, FL presents a strategic opportunity for insurers to innovate and unlock new business.

The process of developing and executing these use cases also revealed several key factors essential for the successful adoption of FL. Strong coordination and clear communication between stakeholders is essential to navigate challenges related to data privacy, model performance, and technical integration. Establishing a partnership agreement can help define data ownership, usage rights, and each party's expected contributions, thereby reducing potential misunderstandings and promoting collaboration. Implementing a comprehensive FL platform that manages the full data lifecycle, from data processing to customised analysis, can improve operational efficiency and encourage broader adoption. Furthermore, ensuring infrastructure is scalable is essential to accommodate increasing data volumes and computational demands.

Looking forward, cross-sector partnerships, clear regulatory support, robust privacy protocols, and technology advancement will be the keys for unlocking the full potential of FL in the insurance sector.

Structure of the Whitepaper

- **Part One: Alternative Data for the Insurance Industry**

  This section examines the challenges faced by the insurance industry regarding data availability and quality. It highlights the need for high-quality data and diverse data sources for accurate predictions, and introduces FL as a potential solution for harnessing third-party data.

- **Part Two: Federated Learning in Insurance: Exploring Risks, Regulations, and Strategies**

  This part introduces what FL is, and explores its potential risks for the insurance sector. It also addresses the issues organisations must consider before adopting FL, emphasising the importance of regulatory compliance and risk management.

- **Part Three: Federated Learning Infrastructure for the Insurance Industry**

  This section proposes a framework for implementing FL in insurance. It discusses the need for an Insurtech infrastructure capable of managing data sourcing, structuring, privacy, and decision-making processes.

- **Part Four: Technical Evaluation of the Proposed Framework**

  This part assesses the technical feasibility of the proposed framework using open-source insurance datasets. The evaluation examines how alternative data and varying data volumes affect the performance of different ML models.

- **Part Five: Proof-of-Concept Work**

  This section describes three use cases developed during the PoC phase, to show some practical applications of the proposed framework across diverse business tasks. It also discusses key considerations for the successful implementation of FL.

- **Part Six: Roadmap for the Future**

  The final part offers a roadmap for the adoption of FL in Hong Kong. It emphasises the need for a multi-pronged strategy, encompassing technical advancements in areas like efficiency, scalability and security, organisational changes to drive its adoption among insurers, and cross-sector collaboration to enhance data availability and collective defence against emerging threats.

Part One

Alternative Data for the Insurance Industry

Part One:
Alternative Data for the Insurance Industry

Data plays a critical role in the insurance industry. Insurance companies rely heavily on data to make informed decisions, assess risks accurately, and provide personalised customer experiences.

Expanding the use of data sources has the potential to help the insurance industry in Hong Kong to address several challenges that could impact its long-term growth and stability. The industry's over-reliance on long-term, non-linked individual life and annuity products with significant saving and investment elements has limited the range of insurance solutions available to consumers, making it vulnerable to changes in customer preferences and market conditions. Additionally, variable underwriting performance in the general insurance market has exposed the industry to potential volatility and external shocks. Leveraging data analytics and emerging technologies could help address these problems, enabling insurers to diversify their product offerings, improve their underwriting, and enhance their risk management.

This part gives an overview of the insurance sector landscape and highlights the current challenges and opportunities, particularly with respect to exploring diverse data sources to enhance the efficacy and efficiency of insurance operations. It suggests the use of federated learning (FL) as a way of insurers broadening the data sources available for making informed business decisions while at the same time protecting data privacy.

1.1 Challenges and Opportunities in the Insurance Industry

1.1.1 Insurance Market Overview for Hong Kong

The Hong Kong insurance market is a well-developed and competitive one that encompasses various segments, which mostly fall into two categories: general insurance and long-term insurance. Table 1 below shows various common types of insurances in these two categories.

The insurance industry is a major driving force in Hong Kong's economy. Despite the challenges posed by COVID-19 and the subdued economic recovery that has followed, Hong Kong's insurance market remains highly advanced and competitive, with impressive insurance density and penetration rates.

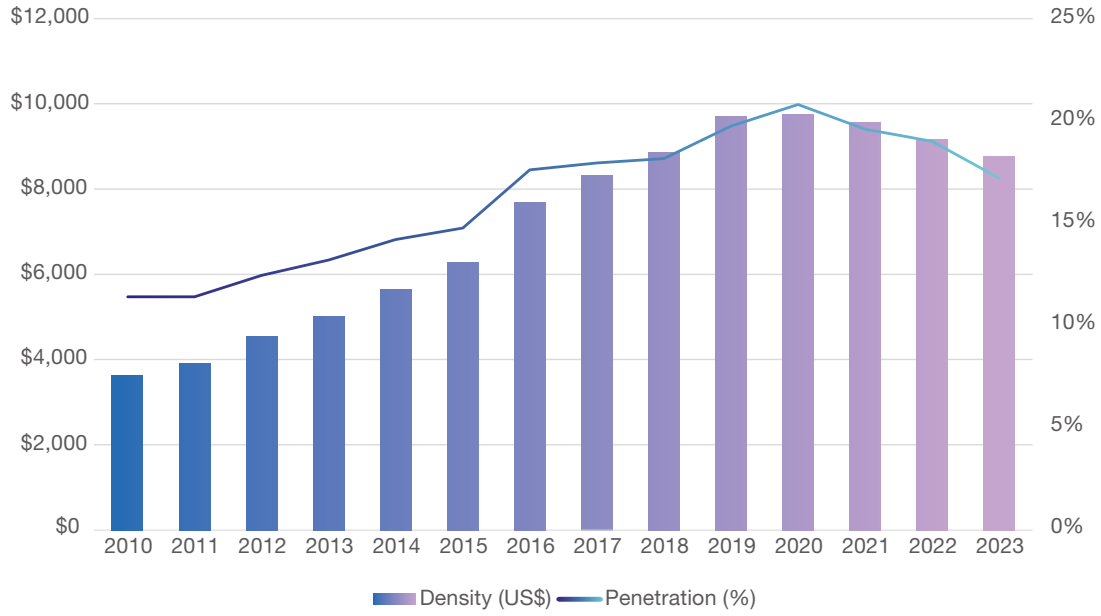
Table 1 Major types of insurance business

Categories	Insurance types	
Long-term insurance	• Life and annuity • Marriage and birth • Linked long term • Permanent health	• Tontines • Capital redemption • Retirement schemes
General insurance	• Accident • Sickness • Vehicles • Fire and natural forces	• Damage to property • Motor vehicle liability • General liability

In 2024, according to the Insurance Authority's provisional statistics, Hong Kong had a total gross premium of HK$637.8 billion. The insurance density[1] (US$8,769 or approximately HK$68,000) and insurance penetration[2] rate (17%) in 2023 placed Hong Kong second and first in the world respectively[3]. Figure 1 illustrates the trends in insurance density and penetration in Hong Kong over the years. In terms of players, Hong Kong has approximately 160 authorized insurers, six of whom are ranked among the top 10 in the world. The intermediaries market is also very strong, with the city having over 120,000 licensed insurance intermediaries.

Long-term insurance business accounts for the majority of the insurance market in Hong Kong. In 2024, the office premiums[5] for in-force long-term business reached HK$537.4 billion. Notably, within the non-linked business[6], with-profits business[7] dominated, accounting for a significant 90.7% of in-force office premiums, suggesting that the market tends to favour products with savings and investment elements. New office premiums were HK$219.8 billion, mainly composed of HK$208.1 billion derived from non-linked individual business and HK$11.2 billion derived from linked individual business. Overall, mainland visitors generated $62.8 billion, accounting for 28.6% of the total individual new business in the same year. Figure 2 shows a more detailed market breakdown of long-term insurance business.

Figure 1
Hong Kong's insurance market density and penetration[4]



1 Insurance density refers to the ratio of insurance premiums to the total population.
2 Insurance penetration refers to the ratio of insurance premiums to GDP of an economy.
3 Swiss Re Institute, World Insurance: Stirred, and Not Shaken, July 2023.
4 Swiss Re Institute and the Insurance Authority.
5 Office premiums in relation to a financial year of an insurer, means: a) for policies with single mode of payment, the premiums paid by policy holders during the financial year; or b) for policies with regular mode of payment, the annualised premiums of the policies as at the valuation date or the flexible premiums paid by the policy holders during the financial year.
6 Non-linked business refers to policies that are not linked to the stock market, meaning that their returns are not based on how the market performs. Linked business, on the other hand, refers to policies that are linked to the stock market, with returns based on how the market performs.
7 With-profits business means business in which policy holders are entitled to participate in the distributable surplus of the insurer, in addition to receiving their contractual benefits.

Figure 2
Market profile of long-term insurance business in 2024

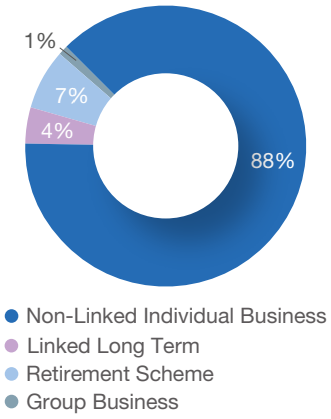
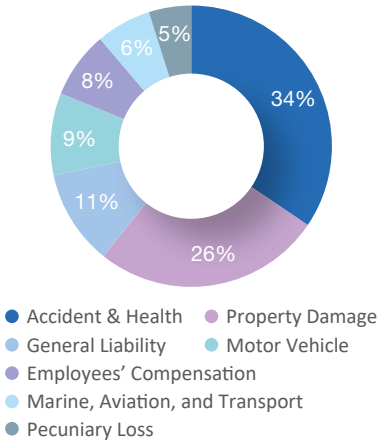


Figure 3
Market profile of general insurance business in 2024



In 2024, including direct and reinsurance businesses, the general insurance industry generated a total of HK$100.5 billion in gross premiums. As shown in Figure 3, the general insurance sector has multiple insurance categories, including property damage, accident and health, and general liability, giving it a more diverse composition. However, the general insurance sector's underwriting performance fluctuates quite significantly from year to year, as it often acts as a shock absorber for society at large. A wide range of factors, such as the extreme weather conditions, can influence the sector's financial results across multiple lines of business. Managing these financial fluctuations effectively is essential for insurers to maintain operational efficiency.

Notwithstanding its considerable market size and relatively developed status, the insurance sector in Hong Kong still has room for improvement. Currently the market is largely driven by the long-term insurance sector, which mainly consists of non-linked individual life and annuity products with saving or investment elements. This indicates a potential opportunity to diversify and expand the range of insurance offerings. Furthermore, the fluctuating underwriting performance of general insurance business suggests there is further room to enhance efficiency and support growth in this sector.

Other variable and unpredictable circumstances, such as extreme weather due to climate change, increasing geopolitical risks and the development of competing insurance markets, further emphasise the need for continuous adaptation and improvement within the Hong Kong insurance industry.

One way for Hong Kong to maintain its comparative advantage in insurance is for Hong Kong insurers to explore ways of leveraging data across the value chain. Data analytics and technology have the potential to provide insurers with important insights into client behaviour, risk patterns, and market trends, thus facilitating more accurate risk assessment, streamlining claims processing, and enabling the development of customised insurance solutions. Ultimately, this should result in improved underwriting performance, enhanced customer experience, and a boost in profitability.

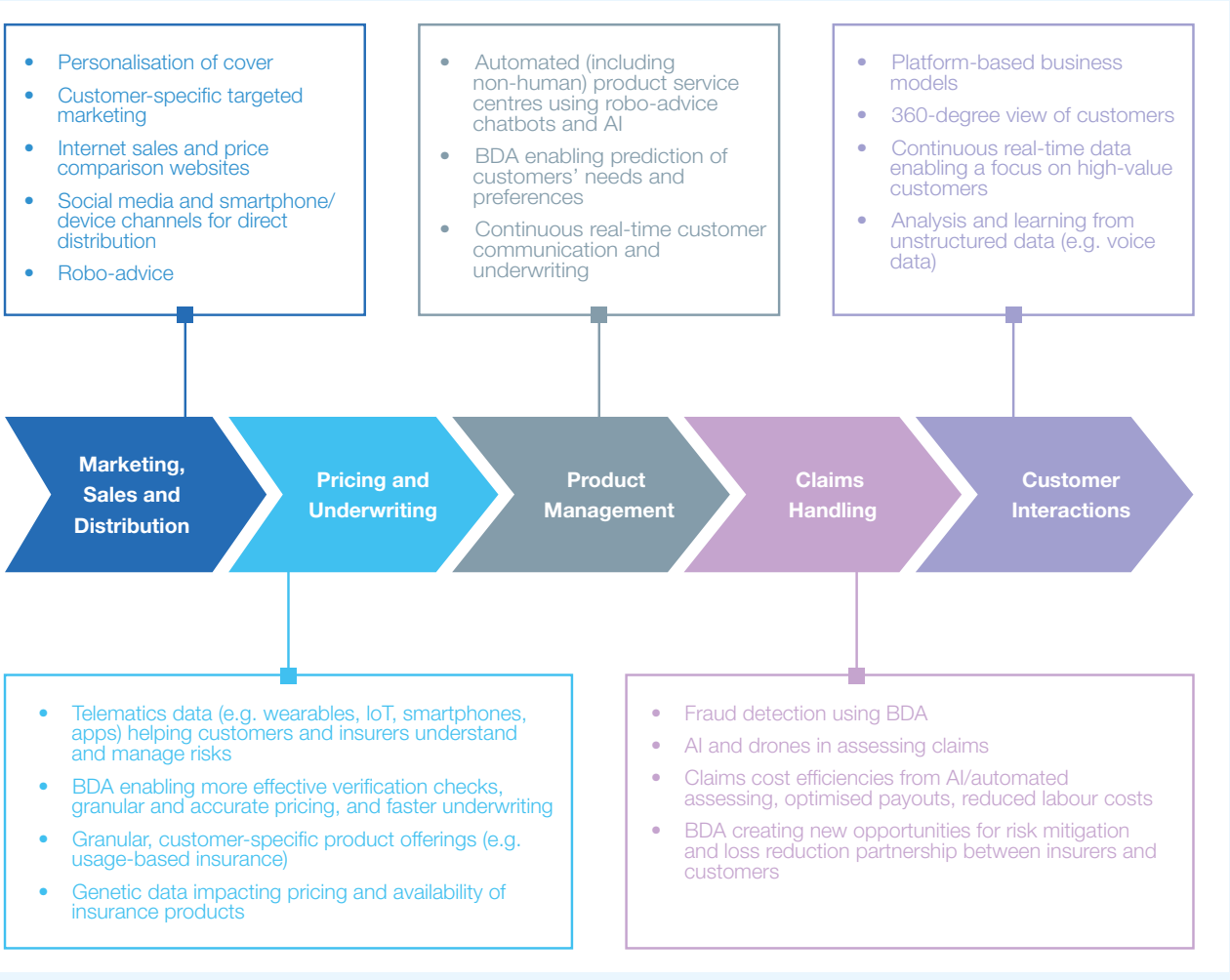
1.1.2 Importance of Data in Insurance

The insurance industry is underpinned by data. Traditionally, insurance firms have relied on internal and structured data, such as demographic information (e.g. age, gender, occupation) and basic health-related particulars provided by customers, to inform their underwriting decisions, determine premiums, assess and settle claims, and combat fraud. In today's era of digitalisation, these traditional datasets are increasingly being combined with new types of data generated and collected from sources outside the company's own operations and databases, such as third-party providers, internet-connected devices, social media platforms, and other external sources, enabling more sophisticated and comprehensive analyses to be undertaken.

Furthermore, more insurance companies are turning to big data analytics (BDA) tools such as artificial intelligence (AI) and machine learning (ML) to enhance their efficiency and reduce their operational costs. According to a 2025 survey conducted by the Insurance Authority, 20% of insurers in Hong Kong have established a strategy to steer AI adoption and are implementing AI applications. Over half are in the exploratory or pilot phases, while 40% plan to expand their AI investments within the next two years. The ability to effectively collect, organise, analyse, and utilise data has become a key differentiator for insurance firms, offering a marked competitive edge.

As illustrated in Figure 4[8], rapid changes are evident throughout the insurance value chain due to digitalisation and BDA. They range from product design, underwriting and pricing to marketing and distribution, claims processing and ongoing customer relationship management.

Figure 4 Usage of big data analytics (BDA) across the insurance product lifecycle



8 International Association of Insurance Supervisors (IAIS), Issues Paper on Use of Big Data Analytics in Insurance, March 2020.

1.1.3 Usage of Traditional Data and its Limitations

Traditional data in insurance refers to data gathered from internal industry sources, such as policy management systems (e.g. application forms), claims databases, actuarial tables, and other established data repositories within insurance companies. This data is usually obtained directly from customers, and is managed by the insurance companies. It is essential for insurance operations as it enables insurers to assess risk, set premiums, and manage policies effectively. Table 2 below illustrates typical examples of traditional data utilised in the Hong Kong insurance industry, classified into general insurance and long-term insurance data types:

Table 2 Typical examples of traditional data utilised in the insurance industry

Data Type	General Insurance	Long-Term Insurance
Basic policy information supplied by customers	• Policyholder details (age, income, gender, occupation, smoking and drinking habits, health status, family medical history, medical records, etc.) • Property characteristics (geographic location, age, size, construction type etc.)	• Policyholder details (age, income, gender, occupation, health status, family history and medical records, etc.)
Historical loss information	• Claim frequency and severity data by business line, geography, and industry • Loss development patterns over time • Catastrophe and disaster loss data	• Mortality experience by age, gender, and cause of death • Morbidity experience for critical illness and disabilities • Lapse and surrender rates
Actuarial data	• Exposure data (total insured values, earned premiums) • Reinsurance cost information • Macroeconomic and industry trend data	• Exposure data (policy counts, sums insured, premium volumes) • Reinsurance cost information • Macroeconomic and industry trend data (interest rates, inflation, etc.)

Different insurance products may draw on different kinds of traditional data due to their unique characteristics and risk profiles. For example, Accident and Health insurance in Hong Kong, such as that provided by the Voluntary Health Insurance Scheme (VHIS), determines premiums based on data about an individual's health risks, which generally varies by age and gender. Other health-related data obtained from the application form, including pre-existing medical conditions, family medical history, and smoking habits, can also help assess the insured individual's health risks.

Property Damage insurance evaluates risks using property characteristics, claims data, and loss history. Pet insurers, for example, use application form data on a pet's breed, age, and medical history to analyse property damage risks and determine coverage, while historical claims provide data on the frequency and severity of incidents.

Vehicle specification data, such as model number and year, is important for automobile insurance since it helps determine a vehicle's value, safety features, and risk of theft or accident. Insurers also use driving licence records to assess risk.

Individual life and annuity insurance coverage and premiums correlate with data regarding the policyholder's age and financial situation. Data relating to financial variables such as income, assets, liabilities, and expenses affect coverage, whereas age data closely corresponds with mortality risk. Pricing relies on actuarial data, such as mortality tables, used to estimate mortality and life expectancy.

While traditional data plays a critical role in insurance operations, it does have the following limitations:

- **Limited scope:** Traditional data sources focus primarily on historical claims and policy data, and may not fully capture emerging risks or changing customer behaviours. This reliance can lead to potential blind spots in risk assessment processes, as the data may not reflect the current risk landscape or customer needs.

- **Retrospective nature:** Traditional data is often retrospective, providing insights only after an incident has occurred. This delay is particularly problematic for risk management and fraud detection, where real-time data is essential for effective decision-making.

- **Data isolation:** Data may be isolated within different departments or systems, hindering a holistic view and making it challenging to integrate insights across the organisation.

Recognising these constraints, insurers are increasingly leveraging alternative data sources to complement traditional data in order to enhance their risk assessment capabilities and make informed decisions more rapidly.

1.1.4 Usage of Alternative Data and its Challenges

1.1.4.1 What is alternative data?

Alternative data refers to information sourced from outside an organisation's databases and operations (including social media platforms, the Internet, wearable and non-wearable sensors, and other external data providers) that can provide valuable insights into the behaviour, preferences, or lifestyle of an entity. It often includes a wide range of unstructured information, including but not limited to social media activity, online shopping behaviour, and sensor data from connected devices such as Internet of Things (IoT) devices.

1.1.4.2 Potential use of alternative data in insurance operations

Alternative data has the potential to revolutionise various sectors of the insurance industry in areas such as product development, customer engagement and interaction, experience monitoring, segmentation analysis and competitor analysis[9]. However, regulations governing its uses in insurance vary significantly among jurisdictions. This section explores several types of alternative data that may be applicable to the Hong Kong insurance industry, based on our desktop research of the global industry landscape. The overview provided below is not exhaustive, and further research on the benefits and challenges associated with leveraging alternative data in the Hong Kong insurance context is necessary.

Like traditional data, alternative data in insurance can be categorised into several broad categories based on data type and sources. Table 3 provides a breakdown of specific categories and some examples of their corresponding sources[10]:

9 Society of Actuaries Research Institute, Alternative Data Usage in Life and Health Insurance: Evidence from Australia, October 2023.
10 Institute of Actuaries of India and India Insurtech Association, Alternate Data Sources in the Insurance Industry, February 2024.

Table 3 Alternative data commonly used in insurance

Data Category	Data Source	Examples	Potential Application and Benefits
Health data	Clinics, hospitals, electronic health records, wearable devices, etc.	• Fitness tracker data • Prescription history • Telemedicine records	**Health Insurance** • More accurate and personalised risk assessment for customised products and premiums. • Early detection of health risks and prevention, possibly leading to improved health outcomes, reduced claims, and stronger customer engagement.
Financial data	Banks, credit rating agencies, online payment platforms, etc.	• Credit card spending patterns • Loan repayment history • Digital payment behaviours	**Life Insurance** • Improved underwriting for individualised life insurance policies tailored to policyholders' financial needs and risk tolerance. • Better longevity risk management • Enhanced customer engagement and retention via personalised financial planning services and advice.
Lifestyle and behaviour data	Online media platforms, fitness apps, search engine providers, IoT devices, telecommunications, etc.	• Exercise routines • Website engagement • App download patterns • Telematics data from connected vehicles • Social media activity • Home sensor data from smart home devices • Shopping habits	**Automobile Insurance** • Better risk management via usage-based insurance (UBI) products. • Accident prevention reduces the likelihood of claims by identifying driving patterns and behaviours associated with higher accident risk.
Geospatial and environmental data	Satellite imagery, weather data providers, property records	• Weather data • Traffic patterns • Neighbourhood characteristics	**Property Insurance** • More accurate property risk assessment. • Improved claim verification, with fewer on-site inspections and faster claims settlement.
Other reference data	Other data sources (e.g. open-source databases)	• Market trends and industry reports • Publicly available statistics • Research studies and academic publications	**Life and Health Insurance** • Better understanding of mortality and longevity risks via the examination of industry reports and market trends on changing lifestyle patterns, medical advancements, and demographic shifts. • Refined risk assessment, benefit design and cost management strategies by analysing public data on disease incidence, healthcare costs, and demographic factors.

• Health data

As defined by the General Data Protection Regulation (GDPR)[11], personal data concerning health encompasses all information that discloses the physical or mental health status of a data subject in the past, present, or future. Aside from the basic health-related particulars provided by prospective policyholders or policyholders at policy inception, insurers may use other sources of health-related data to assess an individual's overall health status.

Substantial amounts of health data are generated by the medical industry, in the form of clinical records, medical images, genomic data, and information on health behaviour[12]. Collaborating with the medical sector and leveraging health data can help insurers more effectively manage health and mortality risks while at the same time enhancing their performance in areas such as modelling, underwriting accuracy, preventive care, claims management, and product innovation.

For instance, analysing the health data from electronic health records (EHRs) and wearable devices such as fitness trackers can help insurers assess risk and customise coverage. In Hong Kong, several life and health insurers[13][14][15] have already deployed points-based wellness rewards programmes through mobile apps and wearable devices. Points-earning opportunities often gather information on an individual's height, weight, physical activity (steps, pace, heart rate), sleep patterns, food choices, blood pressure, cholesterol, and blood glucose. According to case studies from Australia[16], these initiatives can help insurers with their market segmentation efforts by better identifying and targeting healthy individuals. They can also improve customer retention, upselling, cross-selling, and customer modelling by taking advantage of greater customer involvement. In the long run, programme data may enable earlier and more focused health interventions, potentially lowering future claims expenses.

• Financial data

Another valuable alternative data type for the industry is financial data, which reflects an entity's financial condition, transactions, and creditworthiness. Such data includes credit ratings, debt repayment history, transactional data, and other financial metrics obtained from financial institutions. Credit history is often used to underwrite automobile or homeowner's insurance policies. Some insurance companies use their own proprietary formulas to create insurance credit scores based on factors such as payment history, outstanding debt, length of credit history, new credit accounts, and types of credit used.

In Hong Kong, before recommending certain life insurance policies (e.g. annuities), insurers or licensed insurance intermediaries are obliged to conduct a Financial Needs Analysis (FNA), which is a comprehensive assessment that properly assesses the financial circumstances and needs of the customer[17]. Open banking allows third-party financial service providers to access and utilise consumer financial data with permission via application programming interfaces (APIs). In such cases, insurers may use these alternative financial data sources, including records of loan repayment history, credit card spending, investment portfolios, and property ownership, to analyse the customer's financial stability, risk profile, coverage needs, and appropriate premium levels. However, using personal financial data can raise ethical and fairness issues, since such data often includes sensitive information, and individuals who lack access to digital tools or credit cards may be excluded. Furthermore, the Office of the Privacy Commissioner for Personal Data (PCPD)'s Code of Practice on Consumer Credit Data prohibits the use of consumer credit data from a credit reference agency for direct marketing.

11 EU, General Data Protection Regulation (GDPR), 2018.
12 Kornelia Batko and Andrzej Ślęzak, The Use of Big Data Analytics in Healthcare, 2022.
13 AIA, AIA Vitality 健康程式, accessed 5 August 2025, https://www.aia.com.hk/zh-hk/health-and-wellness/aia-vitality.
14 HSBC, Helping Customers Take Steps to Better Health, accessed 5 August 2025, https://www.hsbc.com/news-and-views/news/hsbc-news-archive/helping-customers-take-steps-to-better-health.
15 Manulife, MOVE 計劃及應用程式, accessed 5 August 2025, https://www.manulife.com.hk/zh-hk/individual/products/manulifemove/about-manulifemove/move-program-and-app.html.
16 Society of Actuaries Research Institute, Alternative Data Usage in Life and Health Insurance, October 2023.
17 Insurance Authority (IA), Guideline on Financial Needs Analysis (GL30), September 2019.

• Lifestyle and behaviour data

Many insurance companies are utilising digitalisation to take on a bigger role in their customers' lives. Real-time device trackers that monitor and collect lifestyle and behaviour data are becoming more common in insurance. Data such as social media activity, consumer purchasing patterns, real time device tracker data, and telematics data from connected vehicles is helping insurers better understand consumer habits and preferences.

Social media data can enhance underwriting by providing additional insights and more precise risk assessments. For example, pet-related social media platforms often reveal information about pet owners' engagement with their pets and pet lifestyles, allowing insurers to assess a pet's living conditions and any potential hazards and formulate customised insurance plans.

Lifestyle and behaviour data from fitness trackers and other wearable devices can provide insights into an individual's daily habits, physical activity levels, sleep patterns, and dietary choices. For Accident & Health Insurance, insurers may utilise this information to better understand an individual's lifestyle choices and to promote wellness programmes that encourage healthy behaviours, thus lowering the risk of claims. For Property and Damage Insurance, insurers may use data from IoT devices such as home security systems, smart thermostats, and smoke detectors to gain insights into an individual's lifestyle and living conditions.

Alternative data from telematics devices or mobile applications can help motor insurers accurately analyse driving risks, provide individualised coverage, and promote safe driving practices. Usage-based insurance (UBI) is a prominent example. In the UBI process, telematics devices are put in automobiles to monitor driving behaviour and use, and premiums are determined based on an assessment of the driving behaviour they reveal. Insurance rates are calculated based on a range of parameters, including distance travelled, data time, harsh braking and acceleration, speed, cornering behaviour, and location. Manage-How-You-Drive (MHYD) is one type of UBI that gives drivers real-time feedback that enables them to improve their driving habits, potentially lowering their premiums.

• Geospatial and environmental data

Geospatial and environmental data can enhance insurance risk assessment of specific locations and properties by providing detailed insights from sources such as weather data providers, satellite imagery and property records. Weather data offers historical and real-time information on natural disasters, aiding in accurate risk prediction and timely alerts. Satellite imagery enables precise property assessments, damage evaluations, and risk detection, for example by revealing proximity to flood zones. Property records provide comprehensive details of building characteristics, ownership, and historical claims, crucial for evaluating structural integrity and usage patterns. Integrating these data sources can help insurers improve their fraud detection, and disaster response planning.

1.1.4.3 Potential benefits and issues for the insurance value chain

Using advanced data analytics and alternative data sources in insurance operations has a variety both of potential benefits and challenges, as discussed below[18].

18 IAIS, Issues Paper on Increasing Digitalisation in Insurance and Its Potential Impact on Consumer Outcomes, November 2018.

• Product development and underwriting

Utilising alternative data sources in insurance enables more tailored insurance solutions and product innovations aimed at underserved populations. By leveraging alternative data such as IoT-enabled fitness tracker data (which reveals health and lifestyle choices) or sensor data (which tracks household water usage and identifies potential leaks), insurers can gain deeper insight into their customers, allowing for more customised coverage options and the development of novel preventative or situational insurance products that mitigate risks before they result in significant harm or costly claims.

Additionally, alternative data sources enable underwriting to be based on more granular data, which can improve accuracy and speed up risk-specific underwriting. However, such fine risk categorisation may affect risk pooling principles, potentially leading to affordability issues for certain insurance products and even the exclusion of higher-risk individuals. This could lead to less tech-savvy or less engaged customers being underinsured.

• Risk assessment and pricing

More precise pricing may be possible through the use of alternative data sources. By incorporating additional insights from alternative data, insurers can refine their pricing models and assess risk factors more comprehensively. For instance, financial data can reveal an individual's financial stability or level of financial responsibility, allowing insurers to offer pricing that more closely aligns with the risk profile of the individual policyholder.

However, there are limitations when it comes to pricing and modifying coverage using alternative data. In automobile insurance, variables such as the insured vehicle being driven by someone other than the policyholder may have an impact on data accuracy and premium calculations. It is therefore essential for insurers to ensure that their pricing models are robust and transparent, so that customers can understand how these variables influence their premiums and affect the overall coverage provided. Additionally, customers should be informed whether participation in UBI programmes is compulsory.

Furthermore, adding new data dimensions to long-term insurance products with straightforward premium rates may complicate the premium rate lookup process, making it harder for policyholders to understand the factors driving their rates and potentially eroding consumer trust and engagement. Thus, it may be preferable to adopt a more balanced approach that leverages selected, relevant data dimensions while maintaining a relatively simple and easy-to-navigate overall premium rate structure.

• Marketing and customer experience

The use of alternative data in insurance operations enables insurers to engage in marketing targeted at segments that are more likely to be interested in their offerings, thus enhancing distribution and customer reach while reducing marketing costs.

In Hong Kong, the significant proportion of mainland Chinese Visitor (MCV) business underscores the need for effective cross-boundary data transfer, especially in the context of the Greater Bay Area (GBA) development. Hong Kong insurers need to be able to access and utilise data on mainland Chinese policyholders and their risk profiles in order to provide tailored products and services for this customer segment. Seamless cross-boundary data sharing enables insurers to better understand MCV demographics, behaviours, and risks, and develop insurance solutions that cater to the diverse needs of the GBA market. This access can also enhance customer experience by streamlining the application process, including Know Your Customer (KYC) and underwriting procedures, as well as improving after-sales services for MCVs.

However, potential ethical concerns associated with targeted marketing must be addressed. Unaware of the influence of targeted strategies, customers may end up purchasing products that are not necessarily in their best interests. Insurers should ensure that customers are adequately informed and empowered to make decisions that reflect their actual needs.

1.1.4.4 Common challenges in leveraging alternative data for insurance

As alternative data becomes increasingly prevalent, insurers must navigate the complexities of leveraging this data in an information-rich environment. The following paragraphs highlight a few of the common challenges for the insurance sector in this respect:

• Regulatory compliance and ethical considerations

Hong Kong's Personal Data (Privacy) Ordinance (Cap. 486) (PDPO) establishes a stringent framework for data protection in both the public and private sectors. The Data Protection Principles (DPPs) of the PDPO govern, amongst other things, the collection and use of personal data, and emphasise that data should only be collected for a lawful purpose directly related to a function or activity of the data user, and not to be used for any purpose which is not or is unrelated to the original purpose of collection, except with the express and voluntary consent of the data subject. This can pose challenges for insurers engaging in data exchange, due to strict consent requirements in using the personal data for a new purpose. Concerns about data breaches and security standards can further limit data sharing. Insurers must develop secure data sharing procedures with third-party entities, such as reinsurers or data providers, to ensure compliance with relevant regulations.

There could also be problems with cross-boundary data transfer when insurance companies use alternative data from other jurisdictions. In such transfers, insurers must assess the regulatory implications, ensure strict adherence to relevant regulations, seek expert guidance to navigate the complexities, and implement robust security measures to safeguard data.

Equally importantly, utilising alternative data in insurance requires fairness and transparency. Risk assessment and pricing approaches must be statistically sound and non-discriminatory. Insurers must demonstrate transparency by being able to clearly explain to customers and regulators how the data is being leveraged in their decision-making processes. Failure to uphold these standards can result in regulatory sanctions, reputational damage, and customer backlash.

• Concerns over data quality and quantity

High-quality data is essential for making informed decisions, but insurers may struggle with issues of data quality and quantity when leveraging diverse data sources. One major issue is the prevalence of incomplete or unverifiable data from alternative sources, making it difficult to assess data reliability for critical tasks like risk assessment and underwriting. Robust validation mechanisms and thorough evaluation of the credibility of alternative data providers are necessary in such cases.

Furthermore, the fragmentation of significant amounts of unstructured data in diverse data sources across different systems, formats, and organisations can potentially hinder data integration and interpretation. To overcome these challenges, insurers need to adopt data management practices such as data governance, standardisation, cleansing, and validation. They can also leverage advanced technologies like AI and ML for the purposes of enhancing data accuracy, identifying patterns, extracting valuable insights, and incorporating only the most relevant information from the extensive pool of insurance data.

• Concerns over model security and performance risks

Leveraging alternative data through AI and ML models in insurance raises security and performance concerns. Trustworthy models are essential to prevent bias, errors, and unintended consequences that could affect decision-making. Model explainability and interpretability are crucial for ensuring transparency and regulatory compliance, particularly with complex deep learning models.

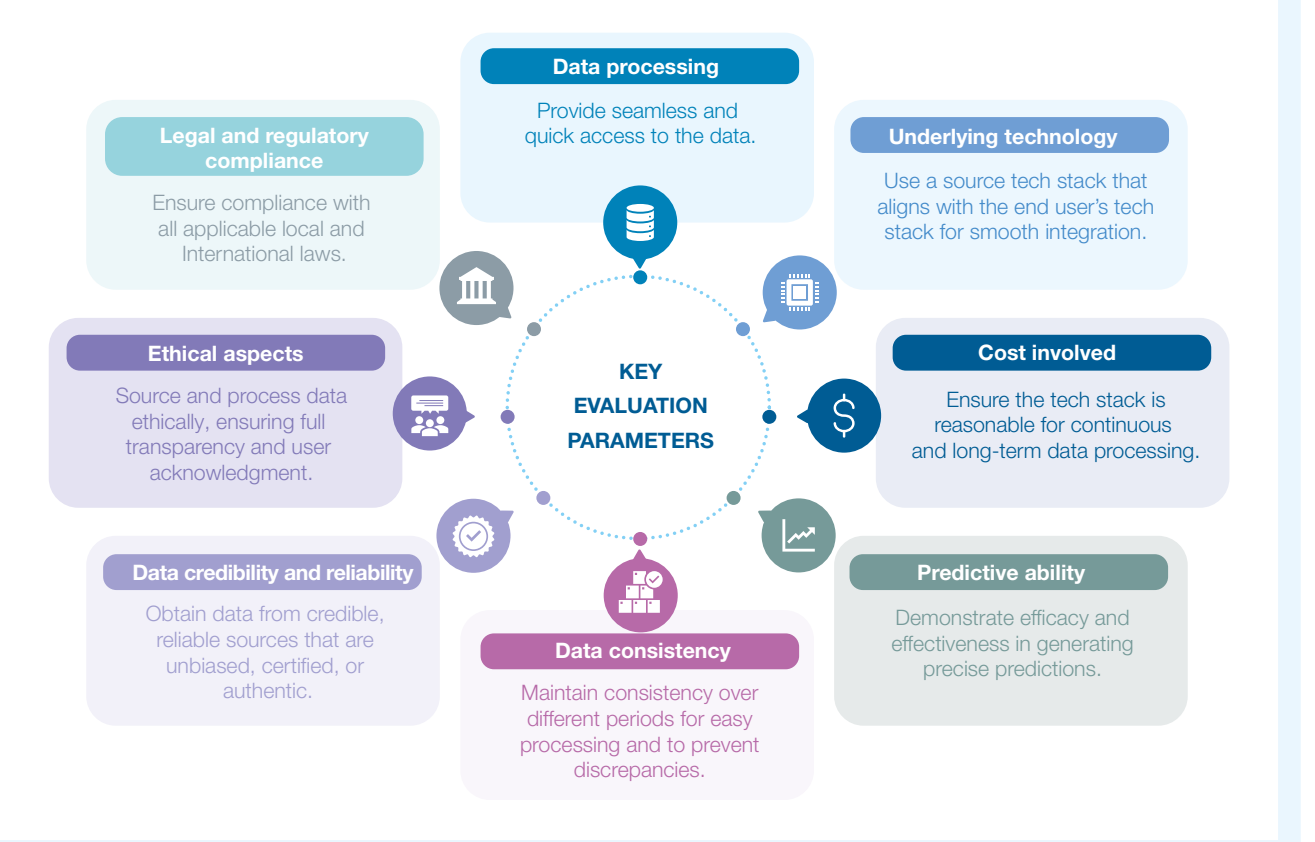
To maintain fairness and mitigate legal and reputational risks, insurance companies must implement robust security measures, such as encryption and access controls, while addressing data biases through preprocessing and regular audits. The integration of large datasets requires substantial computational resources, necessitating investment in scalable infrastructure and possibly cloud-based solutions.

Furthermore, managing data inconsistencies and outliers is critical for enhancing model performance. Rigorous testing, validation, and continuous monitoring will be necessary to detect and manage performance risks, especially since alternative data sources often lack long-term historical data for thorough back testing. Insurers may need to rely on real-time testing and validation, deploying models in controlled environments to monitor their effectiveness and adapt to model drift.

In summary, organisations in the insurance industry must establish clear guidelines for the selection, evaluation, integration, and analysis of alternative data sources to maximise their benefits and mitigate the associated risks. A comprehensive framework should include parameters such as data quality, reliability, relevance, consistency, and ethical considerations.

Figure 5 below illustrates a sample framework of the key evaluation parameters for alternative data sources[19].

Figure 5 Key evaluation parameters for alternative data



1.2 A Potential Solution for Data Security and Privacy: Federated Learning (FL)

ML has been actively explored and implemented in various areas of the insurance industry. However, given the sensitive nature of the data handled and the competitive landscape of the industry, insurers typically approach ML collaboration with a cautious and conservative mindset. To address the challenge of data security and privacy, a potential solution lies in the adoption of federated learning (FL). Hypothetical and real FL use cases, along with their problem-solution-impact analyses, will be presented in Part Four and Part Five.

FL is a revolutionary branch of AI that enables decentralised machine learning, allowing for privacy-preserving data sharing across sectors. Unlike traditional ML approaches that require centralising data in a single location, FL enables models to be trained directly on the devices or servers where the data resides. In traditional ML, clients send raw data to a central

19 Institute of Actuaries of India and India Insurtech Association, Alternate Data Sources, February 2024.

server for model training. By contrast, FL allows clients to send only model update parameters[20] to the central server. This means that participating insurance companies can collaboratively train models without sharing raw customer data, and can access additional data sources to enhance their models and gain new insights by collaborating with other companies and industries. This approach has the potential to significantly enhance insurance model accuracy and efficacy, improving customer experience and business outcomes. The below example illustrates how FL works:

• Example: FL for pneumonia detection

Imagine three hospitals want to build a machine learning model that can detect pneumonia. Each hospital holds private patient records, such as X-rays, lab results, and symptoms, but privacy regulations prevent this data from being shared externally.

To address this, they adopt FL. Instead of sharing raw patient data, each hospital shares only model updates, adjustments to the machine learning model's internal settings that reflect how different medical indicators should be weighted.

The process unfolds in the following steps:

1. **Distribute a base model:** A basic machine learning model is shared with all participating hospitals. It begins with random parameters and must learn which clinical features are most predictive of pneumonia.

2. **Local training with private data:** Each hospital trains the model locally using its own patient data, such as X-rays, lab results, and symptoms. Based on clinical outcomes:

   - **Hospital A** reduces the weight of "cough" as it proves unreliable.
   - **Hospital B** increases the weight of "fever" due to strong correlation.
   - **Hospital C** boosts the importance of "cloudy chest X-ray" as a key indicator.

   These updates are derived from private datasets but do not expose any raw patient information.

3. **Share model updates (mathematical adjustments in weightings and parameters) only without sharing raw patient data:** Hospitals send back only the changes made to the model's internal settings, such as "decrease weight for cough" and "increase weight for fever", without exposing any raw patient data.

4. **Aggregate improvements:** A central server aggregates the updates from all hospitals. Contributions from hospitals with larger datasets or more accurate results may carry more weight in the final model.

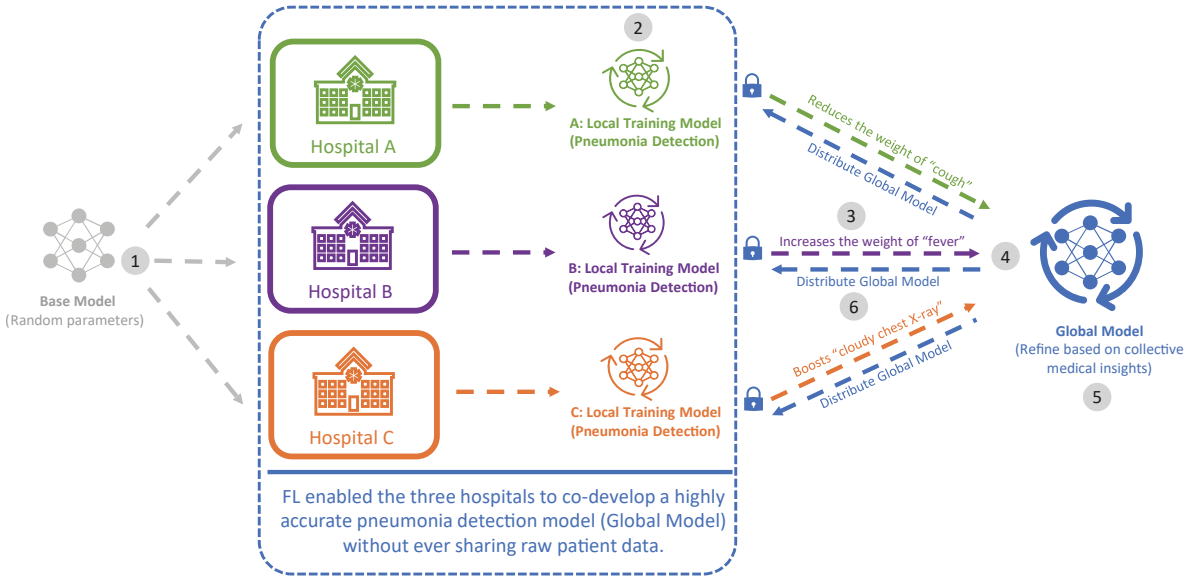
5. **Generate insights without exposing data:** The refined model (Global Model) captures collective medical insights, such as which symptoms are most predictive of pneumonia, without accessing or exposing any hospital's patient data.

6. **Share the improved model:** The Global Model is redistributed to all hospitals, enabling each to benefit from shared intelligence while maintaining full control over their own data.

20 In the context of federated learning (FL), a parameter refers to the numerical values within a machine learning model that are adjusted during training. These parameters, such as weights and biases in neural networks, dictate how the model makes predictions.

As shown in Figure 6, FL enabled the three hospitals to co-develop a highly accurate pneumonia detection model without ever sharing raw patient data. Each hospital trained the base model on its own X-rays, lab results, and symptom records, then contributed only mathematical weight updates. By aggregating these privacy-preserving adjustments, the Global Model captures collective medical insights such as the true predictive power of fever and cloudy chest X-rays while safeguarding patient confidentiality.

Figure 6 Example of Federated Learning for developing a Global Model for pneumonia detection



The use of alternative data sources in the insurance industry often raises significant concerns about data privacy and ethical use, creating a challenge for insurers who need to comply with stringent data privacy regulations like the Personal Data (Privacy) Ordinance (Cap. 486) (PDPO) in Hong Kong and the GDPR. FL offers a solution to this challenge. By enabling model training without the need to centralise sensitive data, it helps insurers address the privacy and compliance concerns associated with alternative data. Furthermore, its scalable and efficient nature can be particularly beneficial for processing and extracting insights from the large, diverse alternative data sources that insurers often work with. Many of these alternative data sources may be too costly or impractical to centralise, making a decentralised approach like FL essential. The key benefit of combining FL and alternative data is the ability to unlock valuable insights and drive innovation in a scalable and efficient manner, while also addressing critical privacy and compliance concerns for the insurance industry.

1.2.1 What Challenges Can FL Potentially Solve?

A typical FL platform enables multiple parties to jointly train a ML model on their decentralised data, thus addressing data security, model security, and technological constraints when utilising alternative data and sharing data with third parties. As shown in Table 4 and elaborated upon in the following sections, an FL platform can potentially solve challenges such as data privacy, data quantity, and model security issues.

Table 4 What challenges can a typical FL platform potentially solve?

A. Regulatory compliance and ethical considerations

Challenges of utilising alternative data	Addressed by a typical FL platform?
1. Data privacy Alternative data sources often contain personal information that raises privacy concerns. Regulatory requirements, such as GDPR and PDPO, impose restrictions on the collection and use of personal data.	**Yes.** FL allows collaborative models to be trained on decentralised data. This preserves data privacy and helps comply with regulatory requirements. Additionally, the FL platform often integrates advanced techniques such as data anonymisation and encryption to enhance data privacy. However, under the PDPO, encrypted data or data that has not yet been fully anonymised may still constitute "personal data", so long as it is reasonably practicable to ascertain the identity of an individual therefrom or when combined with other information held by the data user. Insurance companies should therefore implement a robust set of data protection measures to ensure responsible handling of personal data, including strict access controls and audit trails, and where applicable, obtain explicit, informed consent from customers for the intended use of the data.
2. Cross-boundary data transfers When utilising alternative data in insurance, there are cross-boundary data transfer issues, such as varying data protection regulations, consent requirements, and data localisation rules.	**Partially.** FL localises data and reduces security risks by only exchanging model updates instead of raw data. However, insurance firms must take additional measures to address jurisdictional differences, such as conducting a thorough assessment, ensuring adherence to regulations, seeking expert guidance, and implementing robust security measures.
3. Fairness and transparency There is a risk of bias or discrimination if the data sources or algorithms used to analyse the data are not carefully monitored and regulated.	**Partially.** The FL platform may evaluate the contributions of the training results to assess the fairness of the data sources. Insurance companies should ensure that the data and analytics they use for decision-making clear an extremely high bar in terms of fairness, transparency, and explainability. They have a fundamental obligation to their customers and regulators to demonstrate a solid, unbiased basis for their underwriting decisions. To avoid unfair treatment of customers, it is crucial for insurers to actively monitor and address any potential biases or unfair outcomes that may arise from the platform's analytics.

Table 4 What challenges can a typical FL platform potentially solve?

B. Data quality and quantity

Challenges of utilising alternative data	Addressed by a typical FL platform?
4. Data quality Diverse datasets may come in various formats, making it challenging to integrate and analyse the data effectively. If the alternative data sources have missing data or large variances, they can compromise the output of the model.	**No.** FL does not inherently address all quality challenges, but the platform develops standardised data formats and protocols to facilitate data integration and interoperability. While FL employs data processing and feature engineering techniques such as normalising data, removing outliers, and inputting missing values to enhance data quality, it is important for insurers to prioritise local data preprocessing before engaging in FL to ensure optimal results. By conducting the necessary preprocessing steps, insurers can address specific data quality concerns and improve the overall effectiveness of FL in their operations.
5. Data quantity The cardinal principle of data minimisation emphasises that only a sufficient and relevant amount of personal data should be collected for the intended purpose. This can lead to limited or insufficient data for analysis and decision-making processes.	**Yes.** FL enables distributed data processing and model compression. This allows for efficient analysis and interpretation without the need for centralised data storage and reduces the burden of transferring large amounts of data. Data providers should shoulder the responsibility of implementing the principle of data minimisation while ensuring they provide the efficient and relevant data necessary for accurate model training.
6. Data credibility and reliability Not all alternative data sources provide verified information or disclose their underlying sources, which may result in potentially misleading results.	**Not applicable.** Data providers should bear the responsibility of ensuring that the datasets are credible and reliable. This involves evaluating the reputation, accuracy, and track record of the sources providing the alternative data.
7. Data relevancy Many alternative data sources may not align with insurance-specific use cases. Filtering out non-relevant data can be a challenging exercise for insurance players.	**Not applicable.** The FL platform does not have functions that could filter non-relevant data. Insurance firms should develop effective data preprocessing and filtering techniques to incorporate only relevant and useful data into their analysis. Data providers who fail to provide relevant data may face limitations in their ability to participate as data contributors.

Table 4 What challenges can a typical FL platform potentially solve?

C. Model security and performance risks

Challenges of utilising alternative data	Addressed by a typical FL platform?
8. Model security Model trainings that involve diverse datasets may have a higher risk of unauthorised access or malicious attacks.	**Yes.** The FL platform incorporates robust defence mechanisms to prevent insider threats or back door risks. The database is fully managed by the client at their local/dedicated premises, and data is encrypted.
9. Model performance and efficiency Analysing a large volume of data requires substantial processing power and computational resources. These challenges can impact the overall performance and efficiency of the model process.	**Partially.** FL has the potential to address model performance and efficiency by facilitating the analysis and interpretation of large-scale decentralised datasets and incorporating features to optimise computation time and reduce communication costs. However, there are limitations on the ML algorithms that can be used in the FL context, which may result in suboptimal algorithm choices. Insurers should continuously evaluate the performance of selected ML algorithms within the FL framework for improvement.
10. Lack of historical data for back testing The lack of data archives or historical data in many sources poses challenges for back testing the long-term effectiveness of data usage.	**Not applicable.** To mitigate this risk, insurance companies can explore alternative approaches such as conducting real-time testing and validation of models using current data.

• Enhance data privacy and security

FL trains models locally on devices or servers, significantly reducing privacy risks by keeping sensitive data on-site and sharing only encrypted model updates. This decentralised approach enhances protection against data breaches and unauthorised access, ensuring that data remains under the control of participating organisations. FL aligns with the GDPR's data minimisation principle by keeping raw training data decentralised and preventing unauthorised reuse of personal information. In Hong Kong, the PCPD, in its Guidance on Ethical Development and Use of AI, identifies FL as one of the possible techniques that can minimise the amount of personal data in AI model training by avoiding unnecessary data sharing. However, insurance companies must ensure compliance with relevant regulatory requirements when leveraging FL in their operations. Throughout the FL implementation process, it is recommended that insurance companies undertake a comprehensive review of the regulatory landscape and work closely with legal and compliance teams.

• Overcome data quantity challenges

Insurance companies in Hong Kong may face limitations such as limited bandwidth, network latency, and heterogeneous computing resources when utilising a variety of data sources. A typical FL platform accommodates these constraints by allowing local model training on individual devices or servers. This decentralised approach leverages existing infrastructure and computing resources, enabling participants to train models effectively within their technological limitations. It overcomes data quantity challenges by enabling collaboration and the pooling of diverse datasets without sharing raw data, leading to a larger and more diverse dataset for model training.

- **Mitigate model security and performance risks**

An FL platform may incorporate secure aggregation techniques to combine model updates from multiple participants. These techniques leverage cryptographic protocols to ensure that the aggregated model remains secure and protected during the aggregation process. Additionally, FL enables collaborative model training with diverse datasets, enhancing model performance by capturing a broader range of data patterns and insights. This approach also supports iterative model improvement through continuous collaboration and updates, allowing for ongoing refinement and validation.

1.2.2 Benefits of FL for the Insurance Industry

Despite its promise, it is essential to recognise that FL is still a nascent and developing field. The tangible benefits and long-term return on investment (ROI) associated with its implementation can vary widely based on specific use cases, industry contexts, and the nuances of implementation. Emerging research has documented qualitative advantages and successful use cases of FL across various sectors, including healthcare, financial services, and IoT applications.

As FL has the capacity to address the challenges associated with leveraging diverse data sources, including alternative data sources, it has the ability to unlock the potential of these valuable data assets, translating them into a myriad of benefits for the insurance industry throughout the value chain. The following sections elaborate on some of these benefits.

- **Improved risk assessment**

FL enables insurers to utilise a broader range of data points, leading to more precise risk evaluations. By enhancing insurers' ability to identify potential risks early in the underwriting phase, it allows for proactive mitigation of issues before they escalate. Research has found that using FL can improve loss event prediction by from 30% to 87.5% while also addressing privacy concerns[21].

In addition to identifying patterns across groups of customers, FL also has the potential to be applied to individual risk assessment. This can be done by adapting the global model through fine-tuning, and refining it with individual-specific data. As new data becomes available, the continuous learning of the models can lead to adjustments in the risk assessments for individuals, aligning the model better with an individual's specific situation. The key lies in striking the right balance between the theoretical potential of FL and its practical implementation.

- **Enhanced customer experience**

FL enables insurance companies to leverage insights from various data sources to enhance customer experience throughout the sales process journey, from marketing to post-sales activities.

Prior to sales, FL can harness big data to gain insights into potential customers' behaviour and preferences in order to target marketing campaigns at the right audience, increasing engagement and satisfaction while offering services that are closely aligned with the needs of specific customer groups. It can also strengthen channel relationships by enabling insurers to train data analytics models with their broker partners, helping them to track the status of applications, manage compensation and commissions, and monitor progress towards business goals.

In terms of underwriting, by using FL to analyse data from across different nodes, such as real-time data collected from smartphone apps, insurers can also provide tailor-made products that reflect a customer's unique circumstances on the basis of a better understanding of individual risk profiles and preferences.

Finally, FL boosts data availability by means of its decentralised training on diverse datasets and its real-time updates, which enhance fraud detection and facilitate faster, more reliable settlements.

21 Society of Actuaries Research Institute, Federated Learning for Insurance Companies, February 2024.

• Improved operating efficiency

By utilising FL, insurers can leverage decentralised data sources to automate various tasks without compromising data privacy, thus streamlining traditional operations such as underwriting and claims processing. Automated underwriting processes lead to faster turnaround times, enabling insurers to handle a larger volume of applications in a shorter period. This results in a reduction in operational costs associated with manual labour, paperwork, and data processing.

FL also enhances the efficiency of data analytics. For one thing, it facilitates local data processing, thereby minimising the need for costly centralised infrastructure and reducing the risk of data breaches during transfer. Also, it allows for the continuous improvement of models by using real-time data from personal devices such as smartphones or wearable devices, leading to the development of more accurate and up-to-date models that can adapt to changing market conditions, emerging risks, and evolving customer behaviours.

• Innovation and competitive edge

By harnessing the potential of FL, insurers can unlock novel avenues for product development, driving innovation and staying at the forefront of a dynamic marketplace.

When exploring new insurance products, insurers can collaborate with various stakeholders, such as policyholders, data providers, and even industry partners, without the need to centralise or share sensitive data. By accessing distributed data through FL, insurers can gain a comprehensive understanding of customer preferences, behaviour patterns, and emerging trends.

For instance, insurers can use FL to analyse data from connected devices in the IoT ecosystem that can reveal crucial information about risks associated with smart homes, connected cars, or wearable devices. Armed with these insights, insurers can identify untapped market needs and gaps in their offerings, using these to develop innovative policies that provide coverage against emerging risks, such as cyber threats to smart homes or personalised health insurance plans based on wearable device data.

Part Two

Federated Learning in Insurance:

Exploring Risks, Regulations, and Strategies

Part Two:

Federated Learning in Insurance: Exploring Risks, Regulations, and Strategies

This part first introduces the fundamentals of FL, including its classification, existing frameworks, and applications. It then discusses different aspects of risk management and regulatory compliance relevant to FL implementation. As FL involves multiple stakeholders and sensitive data, it is important to address risks such as data privacy concerns, model accuracy, and security vulnerabilities. Understanding both the technical aspects of FL and the necessary risk management and compliance measures is important in promoting its ethical and responsible implementation.

2.1 What is Federated Learning (FL)?

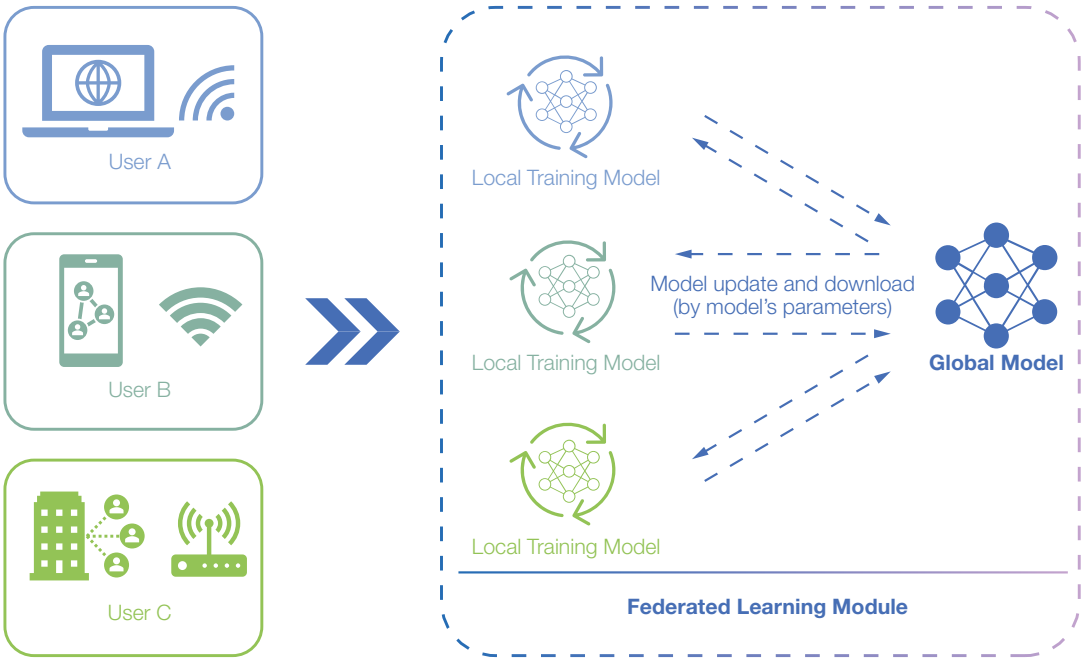
The initial concept of FL can be traced back to a paper published by Google researchers in 2016[22]. It recognised that in many scenarios, data is distributed across multiple devices (edge nodes), with privacy concerns, network limitations, or regulatory constraints making it difficult or impractical to aggregate the data in a central location.

> **Federated Learning enables collaborative machine learning while safeguarding data privacy and security.**

FL presented the novel idea of decentralising the learning process, allowing each individual or organisation (data node) to train a local model using its own data. Instead of sharing raw data, the central server exchanges only the model's updates. Figure 7 provides a visual representation of this process, showing the flow of the exchange of model updates between the local devices and the central server.

22 Jakub Konečný et al., Federated Learning: Strategies for Improving Communication Efficiency, 2016.

Figure 7 Working flow of federated learning



Typically, the FL framework process consists of the following eight steps:

1. **Data localisation:** In an FL setup, data remains on the local devices or servers where it is generated. Each participant, such as a device or organisation, holds its own dataset without transferring it to a central server.

2. **Model initialisation:** A global model is initialised on the central server. This model serves as the starting point for training and is typically based on prior knowledge or a pre-existing model structure.

3. **Local training:** Each participant trains the global model on its local dataset. This involves running multiple iterations of model training using local data while ensuring that data does not leave the device.

4. **Model update collection:** After local training is complete, each participant sends its model updates to the central server, but not the raw data. These updates represent the learned information from the local datasets.

5. **Aggregation of updates:** The central server collects the model updates from all participants and aggregates them to create a new global model.

6. **Model distribution:** The newly aggregated global model is then sent back to the participants. Each participant replaces its local model with the updated global model, which incorporates learnings from all participating devices.

7. **Iteration:** Steps 3 to 6 are repeated for several rounds, allowing the model to improve over time until certain convergence criteria are met, such as a predefined number of iterations or a satisfactory level of model performance.

8. **Final model evaluation:** Once the training process is complete, the final model is evaluated to assess its performance.

2.1.1 Classification of FL

Various approaches to classifying FL are available. The following discussion introduces two classification approaches, based on participant entities and data distribution characteristics.

From the perspective of collaboration among participants, FL can be categorised into two main types: cross-device federated learning (CDFL) and cross-silo federated learning (CSFL). CDFL is often employed in scenarios where data is distributed across individual devices, such as smartphones, tablets, smartwatches, and smart thermostats. CSFL involves collaboration among organisations or institutions that maintain their own data silos, such as hospitals and banks.

From the perspective of the distribution characteristics of data, FL can be classified into three categories: horizontal federated learning (HFL), vertical federated learning (VFL), and federated transfer learning (FTL). HFL pertains to scenarios where multiple data partners collaboratively train a model using the same feature space, meaning the data consists of similar types of information or characteristics, such as standardized tumor images from different hospitals. VFL, on the other hand, involves scenarios where data sources have different types of features regarding the same set of samples, such as the health insurance records and hospital data of the same client. FTL focuses on the transfer of knowledge or models across different FL setups.

To provide a comprehensive overview and facilitate comparison, these two classification approaches have been summarised in Table 5 and Table 6.

Table 5 Comparison of CDFL and CSFL

Aspect	Cross-device federated learning (CDFL)	Cross-silo federated learning (CSFL)
Client entity	Individual devices (e.g. smartphones, wearables)	Organisations or companies (e.g. hospitals, banks)
Data distribution	Generated locally and remains decentralised	
Client scale	A large number (up to a million clients)	A small number (from two to 100 clients)
Bottlenecks	High communication cost and low efficiency	Heterogeneous data (High variability of data types and formats)
Use cases	• Next-word prediction • Personalised recommendations • Health monitoring • IoT applications	• Disease diagnosis and prediction, medical image analysis, drug discovery • Credit risk assessment, fraud detection, market prediction • Smart city development

Table 6 Comparison of HFL, VFL, and FTL

Aspect	Horizontal federated learning (HFL)	Vertical federated learning (VFL)	Federated transfer learning (FTL)
Data distribution	Differ in sample space	Differ in feature space	Differ in both sample and feature spaces
Scenarios	Cross-device/Cross-silo	Cross-silo	Mostly cross-silo
Exchanged items	Model parameters	Intermediate results	Intermediate results

2.1.2 Applications and Emerging Trends

Increasing concerns over customer privacy and data protection have seen companies across various industries recognise the importance of adopting advanced technologies that preserve data privacy while enabling collaborative data analysis. Table 7 below summarises applications of FL in various areas.

Table 7 Application of FL in various areas

Areas	Description	Examples
Finance	FL enables collaboration among financial institutions or online platforms to train robust fraud detection models without sharing sensitive transaction data.	• PayPal has developed a FL platform which allows multiple businesses to train a model that can detect fraudulent transactions without sharing their underlying data[23]. • Amazon has also developed a FL platform allowing multiple e-commerce businesses to train a model that can detect fraudulent orders[24].
Healthcare	Healthcare institutions can use FL to enhance predictive models for disease outcomes, such as predicting patient readmissions or identifying individuals at risk of developing certain conditions. FL can also support personalised treatment recommendations.	• The Clara platform by NVIDIA enables secure collaboration among healthcare institutions for AI model training in medical imaging, genomics and drug discovery[25]. • Project InnerEye by Microsoft develops AI tools for analysing 3D medical images[26]. • The Google Health Studies app utilises FL to facilitate respiratory illness research by collecting user health data while ensuring privacy[27].
Smart Cities	FL can be used for anomaly detection in IoT devices, where the global model learns from the local anomalies detected by each device, improving the overall accuracy of the anomaly detection system.	• Google Maps utilises FL to improve accuracy in predicting traffic congestion and travel times by leveraging decentralised user data, providing real-time updates to users[28]. • In 2023, Bosch and the Austrian Institute of Technology launched a research collaboration to explore the application of FL to a wide range of Bosch products, particularly in the area of Internet of Things (IoT) applications[29].
Natural Language Processing (NLP)	FL can enhance NLP tasks, such as improving the accuracy of sentiment analysis, language translation, and chatbot development, without accessing user data directly. Furthermore, FL-based LLM training frameworks can incorporate additional privacy-enhancing techniques to further strengthen the protection of data privacy and security during the training process.	• Gboard, a keyboard app developed by Google, utilises FL to refine neural network language models for better text prediction and translation accuracy without exporting sensitive user data to servers[30]. The FL environment gives users greater control over their data and simplifies the task of incorporating privacy by default with distributed training and aggregation across a population of client devices.

23 TWIML AI Podcast. Applied AI/ML Research at PayPal with Vidyut Naware, accessed 5 August 2025, https://twimlai.com/podcast/twimlai/applied-ai-ml-research-at-paypal-with-vidyut-naware/.
24 Amazon, Amazon Fraud Detector Detect Online Fraud Faster with Machine Learning, accessed 5 August 2025, https://aws.amazon.com/fraud-detector/.
25 NVIDIA. 2025. NVIDIA Clara: AI-powered Solutions for Healthcare, accessed 5 August 2025, https://www.nvidia.com/en-us/clara/.
26 Microsoft Research, Medical Image Analysis – Project InnerEye, accessed 5 August 2025, https://www.microsoft.com/en-us/research/project/medical-image-analysis/.
27 Jon Morgan and Paul Eastham, Advancing health research with Google Health Studies, December 2020, accessed 5 August 2025, https://blog.google/technology/health/google-health-studies-app/.
28 Eric Miraglia, Privacy that works for everyone, May 2019, accessed 5 August 2025, https://blog.google/technology/safety-security/privacy-everyone-io/.
29 BOSCH, Research Project Federated Learning, July 2023, accessed 5 August 2025, https://www.bosch.com/research/news/federated-learning/.
30 Ziteng Sun, Improving Gboard Language Models via Private Federated Analytics, Google Research Blog, April 2024, accessed 5 August 2025, https://research.google/blog/improving-gboard-language-models-via-private-federated-analytics/.

2.1.3 Existing Open-source Frameworks and Their Limitations

There are several open-source FL frameworks. Popular ones include TensorFlow Federated (TFF), FedML, FATE (Federated AI Technology Enabler), Flower, FederatedScope, FLUTE (Federated Learning Utilities and Tools for Experimentation), and FedScale. While they all share the key features of a FL framework, including client-side training, server-side aggregation and communication, as well as local simulation, they differ in other features, such as types of ML models and libraries supported, ease of customization, privacy protection methods, readiness for real-world use, and compatibility with different devices and operating systems. Table 8 provides a comparison of these frameworks[31].

As FL is a relatively new concept, most frameworks are still under constant development. A framework that demonstrates higher project maturity and offers comprehensive documentation is often viewed as more reliable and better suited for long-term adoption.

Open-source FL frameworks have democratised the development and deployment of FL solutions, providing a foundation for researchers and entrepreneurs to collaborate, experiment, and build on existing FL technology. However, like any emerging technology, open-source FL frameworks come with their own set of challenges, which include:

- **Limited security modules**

When evaluating open-source FL frameworks, it is important to consider their privacy and security features. The origin and nature of these frameworks and platforms vary, with some developed by scientific research projects and others by commercial entities. Not all are supported by professional teams dedicated to security technology, meaning they are vulnerable to potential attacks. Given the importance of data and model protection in FL and the evolving nature of attacks, basic security modules are insufficient and there must be a proactive and comprehensive approach to security design and implementation.

Table 8 Details of FL frameworks

Name	Release Source	Pros	Cons
FATE (2019)	Webank	• Suited for commercial use, with many FL algorithms	• Difficult to extend
TFF (2019)	Google	• Easy to use and flexible	• Limited to TensorFlow/ Keras
Flower (2020)	University of Oxford	• Easy to use and flexible	• Limited extra features
FedScale (2021)	University of Michigan	• Scalable and extensible	• Complex implementation
FedML (2022)	FEDML Nexus AI	• Easy to use and flexible	• Limited performance optimizations
Federated Scope (2022)	Alibaba	• Convenient usage and flexible customization	• High communication costs
FLUTE (2022)	Microsoft	• High-performance FL simulations at scale	• Difficult to extend

31 Alex Braungardt, Flower & PySyft & Co: Federated Learning Frameworks in Python, Medium, 2023, accessed 5 August 2024, https://medium.com/elca-it/flower-pysyft-co-federated-learning-frameworks-in-python-b1a8eda68b0d.

- **Scalability challenges**

Open-source FL frameworks must have a scalability that enables them to bridge the gap between scientific research and practical application. Many open-source frameworks neglect to optimise communication channels for large-scale FL networks, leading to scalability challenges. FL platforms must incorporate robust defence mechanisms and efficient communication protocols. They should also include intelligent mechanisms for upload and download requests, to ensure that resource utilisation is fair and efficient.

- **Lack of module support**

Many open-source FL frameworks face significant challenges due to delays in updates and lack of comprehensive module support. These issues may complicate their integration with existing libraries for ML, deep learning, and Large Language Models (LLMs), such as ChatGPT and DeepSeek. This limitation restricts the ability of FL to fully leverage the potential of available data.

By incorporating the lessons learned from analysing existing FL frameworks, the FL platform proposed in this white paper will integrate privacy-enhancing technologies (PETs) such as differential privacy and secure multi-party computation in order to effectively address privacy and security concerns. A confidential identity matching module (CIMM) will be developed to secure data matching (identity matching or feature matching) across different data sources. The platform also includes a fast-training strategy module (FTSM) designed to enhance training efficiency, thereby lowering scalability costs at the business level. Finally, the platform also focuses on robust modular architectures and provides a variety of algorithms, allowing for easy integration and customisation by different companies. The features of the proposed platform are explained in detail in Part Three of this paper.

2.2 Risk Management and Regulatory Compliance

FL has compelling advantages for preserving data privacy, but it also introduces complexities that demand careful consideration from organisations, making adherence to relevant regulations and mitigation of associated risks imperative. This chapter concentrates on the Hong Kong context, providing local insurers with insights into effectively leveraging FL in their data management practices in ways that comply with the specific risk and regulatory landscape of Hong Kong. This section primarily explores the risks and challenges associated with FL, as well as solutions and mitigation strategies.

2.2.1 Risk Assessment in FL

Common risks associated with FL can be classified into three categories: data privacy risks, model security risks and performance risks. Data privacy risks involve threats to the confidentiality and protection of sensitive data, while model security risks refer to vulnerabilities in the security of FL models. Performance risks relate to issues that can affect the effectiveness and efficiency of the models. Since the application of FL in the insurance industry is still in its nascent stages, it is currently not feasible to assess all the potential risks associated with its implementation.

2.2.1.1 Data privacy risks

During the FL process, data is aggregated across multiple devices or servers. This introduces the risk of data leakage and unintentional exposure of sensitive information. Collaboration between different entities or organisations that own the data sources elevates the risk of unauthorised access by malicious actors within these organisations, potentially leading to misuse of personal or confidential information.

Various solutions can be implemented to address these risks, including secure data storage practices, robust authentication mechanisms, and clear data usage agreements. The following table summarises the risks identified and the associated solutions.

Table 9 Summary of data privacy risks of FL and solutions

Data Privacy Risks	Solutions
Data leakage	• **Secure data storage:** Use secure data centres, encrypt data at rest, have access controls in place, and carry out regular data backups. • **Secure communication channels:** Implement encryption protocols like Transport Layer Security (TLS) for data transmission. • **Data minimisation:** Minimise the amount of sensitive data shared or accessed during the FL process by techniques such as data anonymisation and aggregation.
Unauthorised access	• **Robust authentication:** Implement multi-factor authentication (MFA) to ensure that only authorised individuals can access the data. • **Role-based access controls (RBAC):** Grant access based on job roles and responsibilities. • **Data usage agreements:** Establish clear data usage agreements between participating entities or organisations, outline the permissible use of data, restrictions on data sharing, and protocols for handling and disposing of data after the FL process. • **Data loss prevention:** Deploy data loss prevention solutions to monitor and prevent unauthorised data transmission. • **Monitoring and logging:** Implement robust monitoring and logging systems to track access to sensitive data and analyse logs for suspicious activity. • **Regular access reviews:** Conduct regular access reviews and audits to ensure that access privileges are current and appropriate.

2.2.1.2 Model security risks

An FL framework can be attacked by adversaries, especially if its architecture and parameters are insufficiently protected.

For clients, server trustworthiness may be an issue, as a curious or malicious server could inspect uploaded data and infer private information from it. Expanded client involvement also introduces the potential for malicious actors to manipulate the training process. For example, adversaries can pose as honest clients and introduce erroneous updates to maliciously influence the training model's performance.

When aggregating parameters from clients, there is a risk that a server may leak information during transmission, as communication channels may be vulnerable to eavesdropping by unauthorised entities.

The table below summarises some common types of attack and defence strategies in the context of FL.

Table 10 Summary of common security attacks in FL and defence strategies

Attacks	Description	Defence Strategies
Data poisoning	Injecting misleading data into the training set, e.g. flipping labels	**Anomaly detection:** A proactive strategy that utilizes analytical and statistical methods to identify and filter out malicious occurrences that deviate from expected patterns or activities. **Robust aggregation:** Aimed at mitigating the influence of malicious model updates, serving as a defence against poisoning and backdoor attacks.
Model poisoning	Causing the global model to behave undesirably by manipulating the model's updates	**Robust aggregation:** Also effective here as it helps in mitigating the effects of malicious updates.
Backdoor attack	Inserting a hidden trigger into a trained model enabling attackers to exploit it later by activating backdoor behaviour	**Pruning:** By reducing the model's size through selective neuron removal, this can potentially remove or mitigate the effects of backdoor attacks. **Robust aggregation:** Can help in detecting and neutralizing backdoor attacks in model updates.
Evasion attack	Altering the input samples to deceive the model into producing incorrect outputs	**Anomaly detection:** Can detect unusual inputs that might be attempts to evade the model's normal operation.
Attribute inference attack	Deducing sensitive characteristics of individuals by analysing the outputs or behaviour of a model	**Differential privacy:** Introduces noise to the data, making it hard to infer specific details about individuals. **Multi-party computation:** Ensures privacy by distributing computation, reducing the risk of attribute inference.
Membership inference attack	Deducing specific data points of the training dataset, breaching privacy by revealing if the model 'memorised' particular instances	**Differential privacy:** Helps to mask whether specific data points are included in the dataset. **Homomorphic encryption:** Allows operations on encrypted data, thus protecting the training data from being inferred.
GAN reconstruction attack	Utilising Generative Adversarial Networks to reconstruct sensitive or private data used to train the model	**Differential privacy:** By adding noise, it makes reconstruction of individual data points more difficult. **Homomorphic encryption:** Operations on data can be performed without revealing the data itself, preventing reconstruction.

Attacks	Description	Defence Strategies
Model extraction attack	Extracting the parameters, architecture, or intellectual property of a trained machine learning model to replicate or gain unauthorised access to its functionality	**Homomorphic encryption:** Protects model internals by allowing computations without exposing the model's parameters or structure. **Multi-party computation:** Distributes the model across multiple parties, making it harder to extract the complete model without cooperation from all parties involved.

2.2.1.3 Performance challenges

FL faces several performance challenges that can affect model accuracy and efficiency.

The following table summarises the key performance challenges, associated issues, and potential strategies for improvement.

Table 11 Summary of performance challenges and strategies

Performance Challenges	Issues	Strategies
Data heterogeneity	Variability in data quality and representativeness may bias models.	**Data preprocessing and normalisation:** Standardise datasets and address quality issues, such as by engineering numeric features to capture nonlinear relationships, grouping infrequent categories for high-cardinality variables, creating data dictionaries to document types, units, and scaling (e.g. kilometres vs. miles), and using natural language processing (NLP) to extract insights from unstructured data and convert it into structured formats.
	Disproportionate data contributions lead to inaccurate predictions	**Assessment and contribution feedback:** Adjust learning rates based on data quality.
Communication and computation efficiency	Increased clients and data volume strain bandwidth and increase latency.	**Model optimisation:** Use techniques such as stochastic gradient descent (SGD) and adaptive algorithms.
	High communication costs affect overall performance.	**Communication optimisation:** Apply model compression, quantisation, and differential updates.
Federated optimisation challenges	Traditional algorithms may not be suitable for distributed settings.	**Use specialised techniques:** Implement federated averaging and secure aggregation to handle non-IID data.

2.2.2 Compliance and Regulations

To effectively leverage FL technology while mitigating the above risks, insurers are encouraged to proactively engage with a robust set of regulatory expectations. These include compliance with requirements related to: (i) data protection and privacy, (ii) cybersecurity, (iii) governance and control, (iv) outsourcing risk, and (v) fair treatment of customers. Each of these areas is crucial for the responsible deployment of FL technologies.

In addition to general regulatory requirements, global standards are increasingly referenced as best practice within the insurance sector for managing the risks associated with advanced technologies. The International Association of Insurance Supervisors (IAIS), representing insurance regulators globally, highlights the ongoing relevance of its Insurance Core Principles (ICPs) in managing AI-related risks. Published in July 2025[32], its application paper on AI supervision reiterates that insurers remain responsible for understanding and managing these systems and their outcomes. The paper emphasizes a risk-based and proportional approach, focusing on four key areas of governance and risk management that require particular attention: governance and accountability, robustness, safety and security, transparency and explainability, and fairness, ethics, and redress.

By exercising due diligence and adopting best practices, insurers can effectively leverage FL to safeguard personal data while ensuring alignment with applicable laws and regulations.

The following overview highlights key compliance considerations and is intended as a general guide to potential regulatory implications, rather than a comprehensive legal analysis.

2.2.2.1 Data protection and privacy

FL is a ML and BDA approach that may present privacy-related risks and dangers, including:

- Ubiquitous data collection that may infringe individual privacy
- Probabilistic models that can lead to inadequate reasoning and ambiguity as to whether data is authentic or fake
- The potential for algorithmic discrimination and bias
- Lack of transparency around how data is being used and applied
- Unpredictable or unintended uses of data over time
- Risks of poor data quality and the production of false or misleading information
- Concerns around plagiarism, profiling, and the re-identification of individuals
- Potential for unfair applications and the exploitation of data for wrongdoing

Organisations implementing FL should carefully assess the nature of the data involved and ensure compliance with relevant data protection laws and regulations, including obtaining appropriate consent for its use, anonymising or pseudonymising data when necessary, and implementing security measures to protect the privacy of individuals contributing to the FL process. The following discussion provides an overview of the legal and regulatory landscape surrounding data protection and governance that organisations should pay attention to.

• The Six Data Protection Principles

The Personal Data (Privacy) Ordinance (Cap. 486) (PDPO) is a privacy law in Hong Kong that governs the collection, handling, and use of personal data by both private and public sectors. It sets out six Data Protection Principles (DPPs) that organisations must comply with when handling personal data. The DPPs cover the entire life cycle of personal data in FL, and are summarised in Table 12:

32 IAIS, Application Paper on the supervision of artificial intelligence, July 2025.

Table 12 The Six Data Protection Principles

No.	Principle	Description
1	Collection Purpose and Means	Personal data must be collected in a lawful and fair way, for a purpose directly related to a function/activity of the data user. Where personal data is collected from the data subjects directly, all practicable steps must be taken to notify the data subjects of, amongst others, the purpose of data collection and the classes of persons to whom the data may be transferred. Data collected should be necessary and not excessive.
2	Accuracy and Retention	Practicable steps must be taken to ensure that personal data collected is accurate, and that it is not kept for a period longer than is necessary to fulfil the purpose for which it is used.
3	Use	Personal data must be used for the purpose for which the data is collected or for a directly related purpose, unless voluntary and express consent is obtained from the data subject.
4	Security	A data user must take practical steps to safeguard personal data collected from unauthorised or accidental access, processing, erasure, loss, or use.
5	Openness	A data user must make known to the public its personal data policies and practices, the types of personal data it holds, and how the data is being used.
6	Data Access and Correction	A data subject must be given access to his/her personal data and be able to make corrections if the data is inaccurate.

DPP 4(1) of Schedule 1 of the PDPO requires businesses to take all practicable steps to ensure that any personal data held by them is protected against unauthorised or accidental access, processing, erasure, loss, or use.

While DPP 4 creates an explicit legal requirement regarding the security of personal data, other provisions of the PDPO also have a bearing on data security. Regarding the principle of data minimisation, DPP 1(1) provides that only a necessary and not an excessive amount of personal data should be collected in relation to the purpose for which the data is collected. It is generally accepted that the less amount of data that is collected or held by businesses in the first place, the less exposure to security risks there is likely to be in the future.

On data retention, DPP 2(2) requires a data user to take all practicable steps to ensure that personal data is not kept longer than is necessary for the fulfilment of the purpose (including any directly related purpose) for which the data is or is to be used.

Section 26 of the PDPO provides that a data user is required to take all practicable steps to delete personal data when it is no longer needed for the purpose it was used unless erasure is prohibited by law, or it is in the public interest to retain the data. Implementing data retention policies that ensure the timely deletion of personal data that is no longer needed can help reduce the risk of data breaches.

To ensure compliance with section 26 and DPP 2(2), data users are advised to establish a comprehensive personal data retention policy. This policy should outline the specific retention periods for the personal data they hold. Additionally, data users should develop a personal data erasure policy that provides clear guidelines on management practices for identifying and erasing different types of records, whether in digital or physical format.

Data minimisation, anonymisation, pseudonymisation, deidentification, and timely erasure are some of the possible measures to enhance data protection. Both in theory and in practice, any data on any device is vulnerable to unauthorised or accidental access, processing, erasure, loss, or use.

DPP 2(3) and DPP 4(2) require businesses to adopt contractual or other means to ensure that any data processor engaged by them (e.g. a cloud service provider) also complies with similar requirements in respect of data security and data retention.

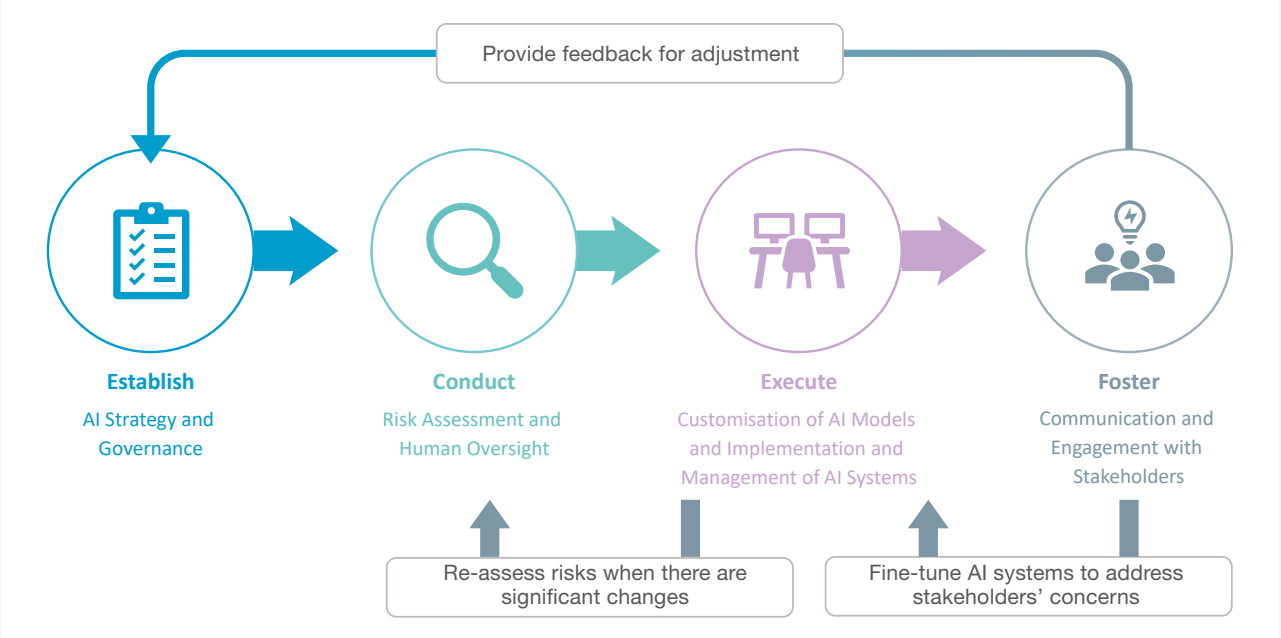
When deciding what "reasonably practicable steps" should be taken to protect personal data, the PCPD would expect businesses to have due regard to the nature of the personal data they hold, the possible impact of a data breach, as well as the technical and organisational measures taken to ensure data security.

If a data security breach is suspected, it is strongly recommended that legal advice is sought as soon as possible. Prompt action in consultation with legal experts can help businesses navigate the complex regulatory environment, mitigate potential damages, and ensure compliance with relevant data protection laws and notification requirements.

• Model Personal Data Protection Framework for AI

In view of the rapid development and wide range of applications of AI (including FL), the PCPD has issued guidelines designed to help Hong Kong enterprises reap the benefits of AI technology while maintaining personal data privacy protection. These guidelines include "Artificial Intelligence: Model Personal Data Protection Framework" (Model Framework)[33] and "Guidance on the Ethical Development and Use of Artificial Intelligence" (Guidance)[34], published in June 2024 and August 2021 respectively. While the Guidance is primarily intended for organisations that develop and use AI systems, the Model Framework targets organisations which procure, implement and use any type of AI systems (including generative AI). Figure 8 below depicts the model personal data protection framework recommended in the Model Framework.

Figure 8 Model Personal Data Protection Framework



33 PCPD, Artificial Intelligence: Model Personal Data Protection Framework, June 2024.
34 PCPD, Guidance on the Ethical Development and Use of Artificial Intelligence, August 2021.

The Model Framework, which is based on general business processes, provides a set of recommendations and best practices for organisations regarding AI governance for the protection of personal data privacy. It is structured to ensure that the governance of AI systems adheres to the three Data Stewardship Values and the Seven Ethical Principles for AI (see Table 13), as advocated in the Guidance of 2021.

Table 13 Data Stewardship Values and Ethical Principles for AI

3 Data Stewardship Values	7 Ethical Principles for AI
1. Being Respectful To respect the dignity, autonomy, rights, interests and reasonable expectations of individuals in processing their data. In this regard, every individual should be treated ethically, rather than as an object or a piece of data.	1. **Accountability:** Organisations should be responsible for what they do and be able to provide sound justifications for their actions. AI-related risks should be assessed and addressed with engagement from senior management and interdisciplinary collaboration. 2. **Human Oversight:** AI system users should be able to take informed and autonomous actions regarding AI systems' recommendations and decisions. When employing AI systems, the level of human involvement should be proportionate to the associated risks and impacts. Human intervention should always be available if the use of AI is deemed high-risk. 3. **Transparency and Interpretability**[35]**:** Organisations should clearly and prominently disclose their use of AI and the relevant data privacy practices while striving to improve the interpretability of automated and AI-assisted decisions. 4. **Data Privacy:** Effective data governance should be put in place to protect individuals' privacy in the development and use of AI.
2. Being Beneficial Emphasises the need to provide benefits to stakeholders, including individuals affected by the use of AI and the wider community, where possible. Meanwhile, any potential harm to stakeholders should be prevented or minimised.	5. **Beneficial AI:** AI should provide benefits to human beings, businesses and the wider community. Provision of benefits encompasses prevention of harm. 6. **Reliability, Robustness and Security:** Organisations should ensure that AI systems operate reliably and as intended over their expected lifetime. AI systems should be resilient against errors during operations, and be protected against attacks such as hacking and data poisoning. Fallback plans should be in place to cope with the failure of AI systems.

35 Interpretability refers to the ability to determine the cause and effect process within an AI system. In other words, it is the extent to which a person can predict what will happen when there is a change in the input to the AI system.

Table 13 Data stewardship values and ethical principles for AI

3 Data Stewardship Values	7 Ethical Principles for AI
3. Being Fair In respect of processes, 'fair' means that decisions are made reasonably and without unjust bias or unlawful discrimination. There should be highly accessible and effective avenues for individuals to seek redress for unfair treatment. In respect of results, 'fair' means individuals in comparable circumstances should be treated similarly. There should be sound reasons for any differential treatments between different individuals or different groups of people.	7. **Fairness:** Individuals are entitled to be treated in a reasonably equal manner, without unjust bias or unlawful discrimination. There should be sound reasons for any differential treatments between different individuals or different groups of people.

When purchasing, implementing or using AI solutions, organisations should take into consideration the recommended measures in the following four areas (see Table 14) to formulate appropriate policies, practices and procedures. This will help ensure that the Data Stewardship Values and the Ethical Principles for AI are implemented.

Table 14 Recommended measures regarding AI data protection

Steps/Areas	Key Recommended Measures
1 Establish AI Strategy and Governance	• Develop an internal AI strategy • Consider governance issues when procuring AI solutions • Establish an internal governance structure (e.g. an AI governance committee) • Provide AI-related training to employees
2 Conduct Risk Assessment and Human Oversight	• Conduct comprehensive risk assessments • Formulate a risk management system • Adopt a "risk-based" management approach • Balance potentially conflicting ethical principles
3 Customise AI Models and Implement and Manage AI Systems	• Ensure data preparation and management processes align with privacy laws and guidelines • Test and validate AI models throughout customisation and implementation • Ensure system security and data security • Carry out continuous monitoring and review of the AI system • Establish an AI Incident Response Plan
4 Communicate and Engage with Stakeholders	• Establish user feedback channels to ensure effective and regular communication and engagement with stakeholders (e.g. internal staff, AI suppliers, individual customers and regulators) • Ensure proper handling of data access and correction requests • Provide explanations for AI-made decisions and output • Disclose the use of AI systems

By adhering to these principles and measures, organisations can ensure that their AI systems are deployed in a responsible, transparent, and accountable manner, and are in compliance with PDPO requirements.

• Proper handling of customers' personal data for the insurance industry

The insurance industry handles a substantial amount of personal and sensitive data, including contact details, financial information, and medical records. Recognising the unique challenges faced by the insurance industry, the PCPD has issued a specific guidance note with practical advice and case studies to assist insurance institutions in complying with the relevant requirements of the PDPO when handling customers' personal data[36].

The guidance note covers a wide range of personal privacy issues. It gives practical tips on the collection of customers' personal data (including medical data and Hong Kong Identity Card numbers), the engagement of private investigators, the collection and use of personal data in direct marketing, the retention of customers' personal data, the use of data for internal training, the access to and handling of personal data by staff and agents, and the handling of data access requests.

Meanwhile, the PDPO (as amended in 2012) requires businesses/individuals intending to use or provide a customer's personal data to others for direct marketing purposes to clearly inform the customer of such an intention and to obtain their consent in prescribed ways. Failure to do so may attract criminal liability. Organisations operating within the insurance industry must therefore maintain awareness of and strict adherence to the legal requirements surrounding the use of customers' personal data for direct marketing[37].

Furthermore, from a corporate risk management perspective, the Insurance Authority's Guideline on Enterprise Risk Management (GL21)[38] contains specific requirements on data governance related to insurance activities, covering:

- **Data relevance and reliability:** ensuring the use of sufficient, reliable, and relevant data in critical insurance processes such as underwriting, pricing, reserving, and reinsurance.
- **Operational risk mitigation:** implementing safeguards against operational risk events, such as data theft, regulatory breaches, sensitive data disclosure, and business disruption caused by data corruption.
- **Data aggregation accuracy:** ensuring the accuracy and reliability of data aggregation processes, which involve consolidating data from various sources for analysis and decision-making.
- **Monitoring and reporting:** establishing an approach and frequency for monitoring and reporting data quality deficiencies, allowing for timely identification and resolution of issues.
- **Regular review:** conducting regular reviews of data quality controls, systems, and policies to ensure their effectiveness and alignment with industry standards and best practices.

Adhering to these best practices and regulatory requirements will enable insurance institutions to enhance their compliance efforts and safeguard the privacy of customers' personal data throughout their operations.

• Cross-boundary data transfer

Given the increased amount of data collaboration between Chinese Mainland and Hong Kong, Hong Kong-based companies that implement FL across regions should closely consider the various cross-boundary data transfer laws, regulations, measures, and guidelines in Chinese Mainland and Hong Kong.

For transfers of personal data to places outside Hong Kong (including northbound data transfers from Hong Kong to Chinese Mainland), the DPPs under the PDPO apply, regardless of the destination of the data transfer. The PCPD has recommended the use of Recommended Model Contractual Clauses (RMCs) to facilitate compliance with the PDPO's DPPs for cross- boundary data transfers[39]. RMCs set out the general obligations of the contracting parties in respect of the protection of personal data privacy, and cater for two different scenarios in cross-boundary transfers, namely, (i) from a data user to another data user; and (ii) from a data user to a data processor. They are applicable to:

36 PCPD, Guidance on the Proper Handling of Customers' Personal Data for the Insurance Industry, November 2012.
37 PCPD, Guidance on Direct Marketing, April 2023.
38 IA, Guideline on the Use of Internet for Insurance Activities (GL 21), July 2019.
39 PCPD, Guidance on Recommended Model Contractual Clauses for Cross-border Transfer of Personal Data, May 2022.

a. The transfer of personal data from an HKSAR entity to another entity outside the HKSAR, including Chinese Mainland; or

b. The transfers between two entities outside the HKSAR when the transfer is controlled by an HKSAR data user.

Since the signing of the "Memorandum of Understanding on Facilitating Cross-boundary Data Flow within the Guangdong-Hong Kong-Macao Greater Bay Area"[40] on 29 June 2023 between the Cyberspace Administration of China and the HKSAR Government's Innovation, Technology and Industry Bureau, there has been a key development regarding data transfer rules within the Greater Bay Area (GBA):

On 13 December 2023, the "Implementation Guidelines on the Standard Contract for Cross-boundary Flow of Personal Information Within the Guangdong-Hong Kong-Macao Greater Bay Area (Mainland, Hong Kong)" (粵港澳大灣區 (內地、香港) 個人信息跨境流動標準合同實施指引) came into effect[41]. Based on the relevant data protection laws of Chinese Mainland and Hong Kong, these guidelines aim to promote the safe and orderly cross-boundary flow of personal information within the GBA. With effect from 1 November 2024, the facilitation measures of the GBA Standard Contract, piloted in the banking, credit referencing and healthcare sectors, have been extended to cover all sectors in Hong Kong.

• Use of sensitive personal information

Some data used by the insurance sector is considered sensitive in nature, requiring more cautious handling and adherence to specific regulations in Hong Kong or other jurisdictions.

Under the laws of the mainland, sensitive personal information is subject to strict processing rules, and separate or written consent may be required for the processing of such data. PIPL defines sensitive personal information as "personal information that, if leaked or illegally used, may easily lead to infringement of a natural person's personal dignity or endanger the personal safety or the property of a person", including information relating to:

- Biometrics
- Religious beliefs
- Specific identities
- Healthcare
- Financial accounts
- A person's whereabouts
- Any personal information of minors under the age of 14

Unlike in the European Union (EU) or Chinese Mainland, Hong Kong's PDPO does not have a similarly defined classification of "sensitive personal data". Hong Kong companies must therefore remain vigilant and be aware of the differences between Hong Kong's regulations and the more stringent requirements of places like Chinese Mainland and the EU. Given the sensitive nature of the alternative data used in insurance operations, such as medical and financial data, it is recommended that insurance companies in Hong Kong should adhere to very rigorous data protection measures and ethical practices when handling such data.

While the PDPO in Hong Kong does not define "sensitive personal data", the PCPD has provided specific guidance on the collection, use and retention of personal identifiers and consumer credit data through two codes of practice:

40 Digital Policy Office (DPO) (formerly known as the Office of the Government Chief Information Officer, OGCIO), Facilitating Cross-boundary Data Flow within the Greater Bay Area, accessed 5 August 2025.

41 國家互聯網信息辦公室 & 香港特區政府創新科技及工業局, 粵港澳大灣區(內地、香港)個人信息跨境流動標準合同實施指引, 2023年12月13日.

1. Code of Practice on the Identity Card Number and Other Personal Identifiers (Revised in April 2016)[42]

Hong Kong Identity (HKID) Card numbers are commonly collected and used by organisations such as insurers to identify individuals and manage records related to them. However, the indiscriminate collection and improper handling of HKID Card numbers and copies may unduly infringe the privacy of the individuals and create opportunities for fraud. The Code provides guidance on the appropriate handling of personal identifiers in general, and HKID Card numbers and copies in particular. These include:

- Organisations in Hong Kong are required to carefully consider less privacy-intrusive alternatives and give the individual the option of choosing such alternatives before deciding to record or collect an individual's HKID Card number.

- Where an organisation has collected an HKID Card number for a permitted purpose under the Code, they should generally only use that number for that purpose or other further purposes allowed by the Code, and not for any other unauthorised purposes.

- An organisation should not keep records of HKID Card numbers for longer than is necessary to fulfil the purpose for which they were collected.

Insurance providers in Hong Kong must strictly adhere to the requirements of the Code if using customers' HKID Card numbers for identity matching or verification purposes, such as in ML models including FL.

2. Code of Practice on Consumer Credit Data (Revised in January 2013)[43]

This Code is designed to provide practical guidance to data users in Hong Kong for the handling of consumer credit data. It deals with the collection, accuracy, use, security and access and correction issues as they relate to personal data of individuals who are, or have been, applicants for consumer credit. The Code covers, on the one hand, credit reference agencies (CRAs), and on the other hand, credit providers in their dealing with CRAs and debt collection agencies.

As the Proof-of-Concept (PoC) of this white paper research involves the use of consumer credit data held by one of the CRAs, the handling of this data must adhere to the requirements of the Code. These requirements include:

- Credit providers, such as banks and money lenders, are prohibited from accessing consumer credit data held by CRAs for direct marketing purposes. This includes offering or advertising goods, facilities, and services to individuals. However, it does not prohibit a credit provider from accessing the credit data of its existing customers in the course of reviewing or renewing their credit facilities.

- A CRA may not transfer consumer credit data held by it to a place outside Hong Kong unless the purpose of use of the transferred data is the same as or directly related to the original purpose of its collection.

42 PCPD, Code of Practice on the Identity Card Number and Other Personal Identifiers, April 2016 (First Revision). Note: An updated explanatory note titled "Code of Practice on the Identity Card Number and Other Personal Identifiers: Compliance Guide for Data Users" was issued (Revised in August 2024).

43 PCPD, Understanding the Code of Practice on Consumer Credit Data Frequently Asked Questions on the Sharing of Mortgage Data for Credit Assessment Purpose, October 2015.

• Best practice in the use of genetic test results

Many countries have introduced limitations (via self-regulation or legislation) in recent years on requests for and the use of genetic test results by insurers for the purpose of assessing insurance applications.

In Hong Kong, the Hong Kong Federation of Insurers (HKFI) established the Code of Practice on Genetic Testing in 2000, revised in 2020. It sets out key principles and best practices for the use of genetic test results in the insurance sector, which include but are not limited to underwriting and claims assessment[44]. According to the Code:

- Insurers will not require, compel, or pressure potential applicants to undertake genetic testing for underwriting purposes.

- In any event, insurers will not ask for the results of any types of genetic tests (Diagnostic or Predictive) for the purpose of underwriting if the genetic testing was conducted in the context of scientific research.

- Insurers will not ask for or use the results of any genetic tests of a relative or family member of a proposed or existing insured person for the purpose of underwriting.

- Insurers may ask for certain predictive genetic test results only when the applicant applies for Life Insurance or Critical Illness/Dread Disease policies over defined protection limits, e.g. HK$5M and HK$1M respectively. For medical indemnity insurance, no predictive genetic test results will be requested, regardless of the sum insured.

Although the Code is not legally binding, insurers in Hong Kong are advised to adhere to it in order to promote responsible and ethical practices around the use of genomic data.

Insurers should also be aware that genetic privacy is protected to varying degrees across many jurisdictions. In the US, the Genetic Information Nondiscrimination Act of 2008 (GINA)[45] is a federal law that prohibits the use of genetic information in health insurance underwriting and employment decisions. However, GINA does not apply to life insurance, disability insurance, or long-term care insurance, for which the use of genetic information is still largely unregulated at the federal level. In the UK, the Association of British Insurers has also developed a voluntary code of practice[46] that limits the use of genetic test results in insurance underwriting. In the EU, Article 9[47] of the GDPR considers genetic data as a special category of personal data, and its use is subject to strict regulations and safeguards. Recital 52 of the GDPR[48] provides exceptions to the prohibition on processing special categories of personal data, such as for health purposes, public interest, or the establishment, exercise or defence of legal claims.

44 The Hong Kong Federation of Insurers, Best Practice on Use of Genetic Test Results, May 2020.
45 U.S. Department of Health and Human Services, The Genetic Information Nondiscrimination Act of 2008, 2009.
46 UK Government and the Association of British Insurers, Code on Genetic Testing and Insurance, October 2018.
47 EU, General Data Protection Regulation, Processing of Special Categories of Personal Data, 2016.
48 EU, General Data Protection Regulation, Recital 52 Exceptions to the Prohibition on Processing Special Categories of Personal Data, 2016.

2.2.2.2 Cybersecurity

Like any technology that involves data and communication, FL may pose cyber hazards. To prevent and mitigate such risks in the insurance sector, the IA has issued the GL20 Guideline on Cybersecurity, which outlines the minimum cybersecurity standards that authorized insurers in Hong Kong must adhere to in order to safeguard their business data and the personal data of their policyholders.

According to GL20, insurers are required to develop a customised cybersecurity strategy and framework that aligns with the nature, scale, and complexity of their business. The board of directors should hold overall responsibility for cybersecurity controls, and establish a designated management team to oversee and implement cybersecurity measures. A self-assessment tool and systematic monitoring process should also be implemented for overall cyber risk management. Any cyber incident detected must be reported to the IA within 72 hours[49].

The GL21 Guideline on Enterprise Risk Management supplements GL20 by highlighting the importance of incorporating a risk management policy on cyber risk, including controls relating to:

- protecting policyholder data and digital/electronic data
- identifying, preventing, detecting, and mitigating cybersecurity threats
- monitoring and reporting cyber risks
- regular testing of mitigation measures
- communicating cybersecurity policies and procedures to staff, and regularly reviewing and assessing the policies and procedures and monitoring their implementation

2.2.2.3 Outsourcing risk

Developing a FL model in collaboration with external entities or with the assistance of third-party service providers typically exposes organisations to greater operational risks. While Hong Kong does not have any specific statutes that govern and regulate outsourcing arrangements, some relevant industry-specific regulations and guidelines apply.

A major concern when outsourcing IT and cloud services is data privacy. The PCPD advises organisations adopting cloud computing services to fully assess the benefits and risks, recognise the shared responsibility between the organisations as data users and cloud service providers to safeguard personal data privacy, especially data security, in a cloud environment, and ensure they are compliant with the PDPO[50].

In the insurance sector, the IA has issued the Guideline on Outsourcing (GL14)[51] to regulate the outsourcing activities of authorized insurers. In accordance with GL14, an authorized insurer should conduct due diligence in selecting its service provider and ensure its outsourcing arrangements comply with relevant laws and statutory requirements on customer information confidentiality (e.g. the PDPO). GL14 also requires insurers to conduct a comprehensive risk assessment of their outsourcing arrangements, and to put in place a contingency plan to ensure that their business will not be disrupted as a result of undesired contingencies (e.g. system failure) of the service provider. The Guideline also emphasises that the board of directors and management of authorized institutions should retain ultimate accountability for any outsourced activity.

2.2.2.4 Fair treatment of customers

Treating customers fairly is the focus of ICP 19[52] issued by the IAIS. This principle lies at the very core of insurance regulation, as set out in the Insurance Ordinance and reinforced by various guidelines, including the Guideline on the Corporate Governance of Authorized Insurers (GL10)[53], the Guideline on

49 IA, Guideline on cybersecurity (GL20), 2019.
50 PCPD, Guidance on Cloud Computing (Second Revision), January 2025.
51 IA, Guideline on outsourcing (GL14), 2017.
52 IAIS, ICP 19 Conduct of Business, accessed 7 August 2025, https://www.iais.org/icp-online-tool/13530-icp-19-conduct-of-business/.
53 IA, Guideline on the Corporate Governance of Authorized Insurers (GL10), 2017.

Underwriting Class C Business (GL15)[54] , and the Guideline on Underwriting Long Term Insurance Business (other than Class C Business) (GL16)[55].

The use of FL in the insurance industry may introduce considerations related to customer fairness, particularly concerning bias in decision-making and transparency. To manage these risks and adhere to regulatory expectations, insurers should implement the following:

- **Provide adequate and clear information:** Ensure customers receive accurate, timely, and comprehensible information throughout the insurance lifecycle. This includes 1) presenting product information in plain language and bilingual formats, avoiding technical or industry jargon, 2) clearly disclosing key product features and risks, 3) explaining clearly how customer data is used, and how decisions are made, particularly when FL is involved in product recommendations, underwriting or claims. These measures help customers make informed decisions and manage their expectations effectively.

- **Conduct suitability assessments with human oversight:** Before recommending products, insurers must assess their alignment with the customer's needs, financial status, and risk appetite. While FL can assist in these assessments, human oversight is essential to validate model outputs and ensure recommendations serve customers' interests.

- **Give proper advice:** Any advice provided must prioritise the customer's best interest, supported by clear reasoning and documentation. Employees and intermediaries should be equipped with the necessary training to understand both the benefits and limitations of FL, and be prepared to act with skill, care, and diligence.

Additionally, ongoing monitoring and auditing of the FI models are vital, along with maintaining human oversight throughout the model lifecycle to ensure trustworthiness. Insurers should work closely with their technology partners to establish robust model governance and auditability processes.

2.2.3 Recommendations and Conclusion

This chapter has discussed the major legal issues related to FL, including risks, compliance issues, and ethical principles associated with this technology. Insurers are advised to consider the following recommendations before adopting FL.

Firstly, insurers must ensure their technical readiness for FL adoption by conducting a careful examination of the associated risks and establishing robust risk management strategies. This involves thoroughly assessing technological infrastructure, compatibility with existing systems, and potential vulnerabilities. By addressing these considerations upfront, insurers can proactively mitigate risks and ensure a smooth integration of FL into their operations.

Secondly, as the insurance business involves handling vast amounts of sensitive customer data, data protection and security should be a priority in the early stages of project planning. Insurers must diligently comply with relevant data privacy laws and regulations, and implement comprehensive security measures to safeguard sensitive data. They should also closely follow the evolving landscape of data privacy laws in other jurisdictions.

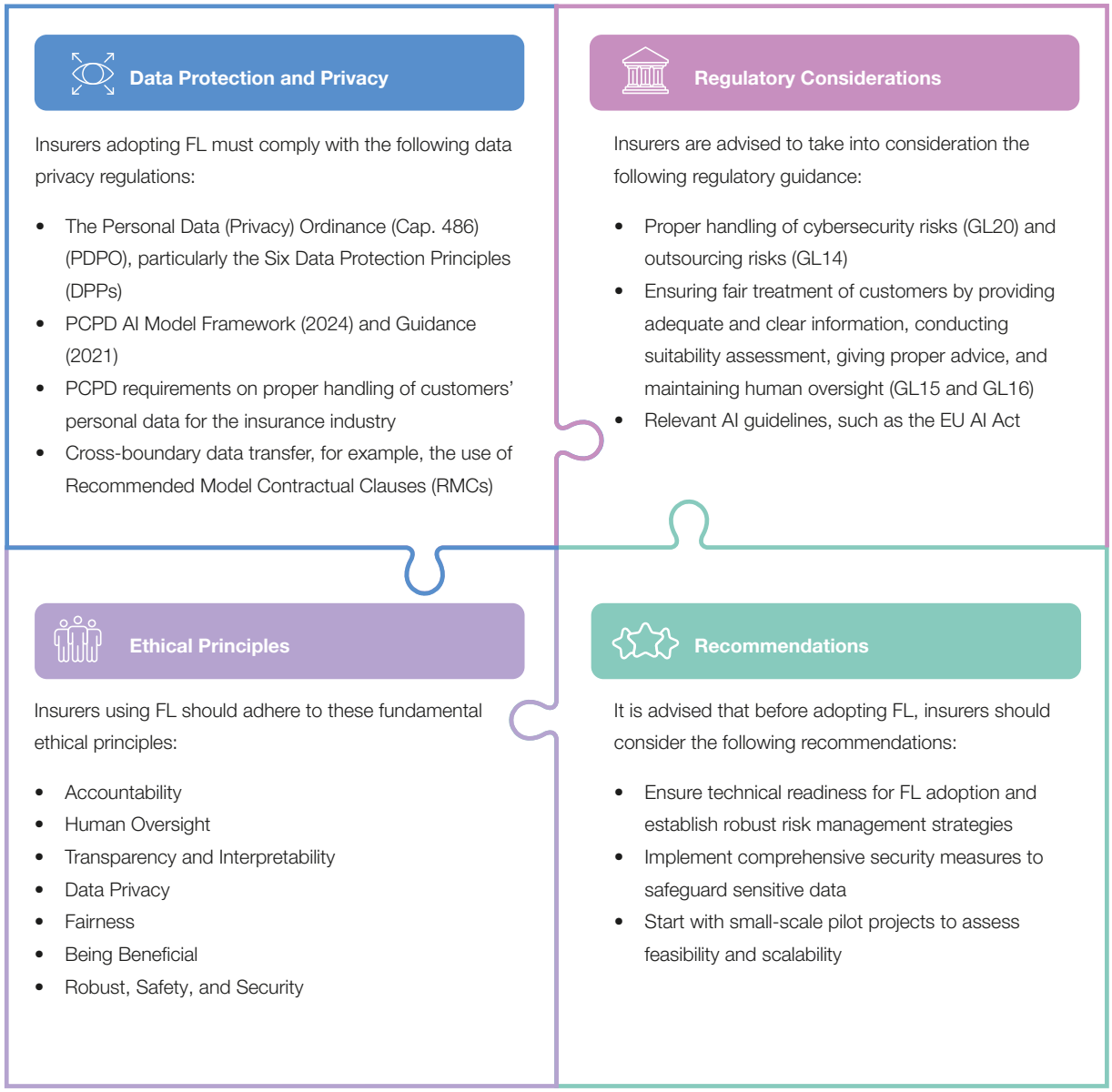
Lastly, given that the application of FL in the insurance sector is still in its early stages, it is recommended that insurers should start with small-scale pilot projects to assess their feasibility and scalability. By initiating pilot projects, insurers can gain valuable insights, learn from initial experiences, and make necessary adjustments before extending FL implementation across broader insurance processes. This iterative approach will also enable insurers to identify potential challenges, fine-tune their strategies, and optimise the benefits of FL technology for their specific business needs.

54 IA, Guideline on Underwriting Class C Business (GL15), 2017.
55 IA, Guideline on Underwriting Long Term Insurance Business (other than Class C Business) (GL16), 2023.

As FL applications become more specialised, it is anticipated that domain-specific legal issues will emerge that will require appropriate resolutions. Therefore, engagement with legal professionals at the earliest stages of any FL-related project is strongly recommended, to ensure that all legal implications are comprehensively considered and incorporated into the project's design. Moreover, given the evolving legal landscape surrounding FL, insurers must stay proactively informed about emerging legal requirements and best practices to ensure their ongoing compliance and mitigate potential risks.

The successful deployment of FL within the insurance industry requires a comprehensive assessment framework that addresses the key considerations and challenges associated with this collaborative ML approach, summarised in Figure 9. This framework should act as a guide for insurers on their journey to leverage FL for various insurance-related applications.

Figure 9 Assessment framework for federated learning in the insurance sector



Part Three

Federated Learning Infrastructure for the Insurance Industry

Part Three:

Federated Learning Infrastructure for the Insurance Industry

This part describes the key infrastructure required to create a collaborative Federated Learning (FL) data analytics platform for the insurance sector. This FL platform is applicable to insurance business scenarios such as product development, risk assessment, claims management, renewal prediction, and fraud detection, areas where collaborative data analysis and modelling can drive innovations while at the same time preserving data privacy. To assist insurers in its implementation, this part also explains how the infrastructure works, and provides an overview of some key Machine Learning (ML) models relevant to the insurance sector. In addition, it discusses mainstream privacy-enhancing techniques, and highlights key advances in areas such as training efficiency and secure identity matching among entities.

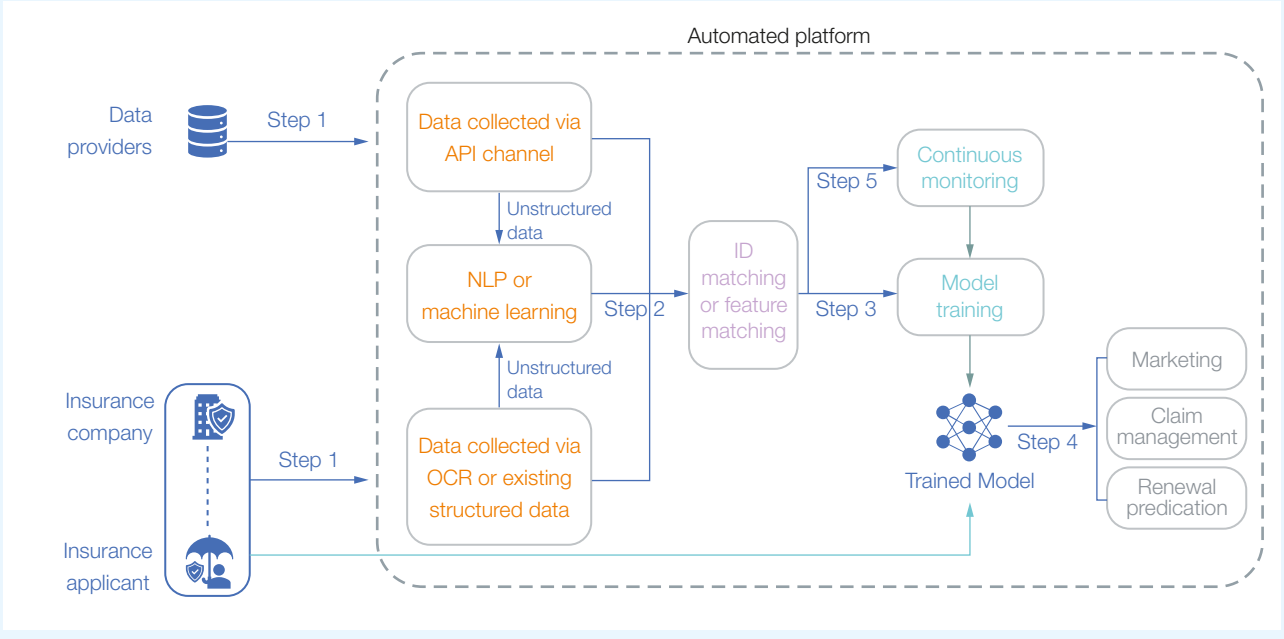
> **Federated Learning redefines collaboration in insurance, enabling organisations to jointly unlock insights from distributed data, driving innovation without compromising privacy.**

3.1 The Federated Learning Collaborative Data Analytics Platform

An automated FL platform collects and structures data from different parties or channels and applies ML models to predict outcomes. It streamlines data collection and analysis by leveraging digital technologies like optical character recognition (OCR) and natural language processing (NLP) to extract information directly from databases or scanned documents and analyse it. For the insurance sector, an FL data analytics platform created through cross-organisation collaboration has the potential to enhance a wide variety of insurance tasks.

The blue arrows in Figure 10 show the key steps in the FL data analysis process. The green arrow indicates how new insurance applicant data can be fed directly into the trained model to generate model decisions in areas such as marketing strategies, claims management, and renewal prediction. This section goes on to describe each step in detail.

Figure 10 The FL data analysis process



3.1.1 Step 1: Decentralised Data Collection

Step 1 is primarily preparation work prior to the actual FL training, focused on organising and standardising the decentralised data collected. In an FL framework, decentralised data is collected from various sources but remains distributed across different organisations, allowing each to retain control over its own data. To optimise this process and ensure the effective integration of a diverse range of data, three key methods can be leveraged: Open APIs, OCR, and NLP.

- **Open APIs** provide a well-defined interface for the secure and standardised exchange of data between the platform and external systems, promoting interoperability and facilitating the integration of diverse data sources. The banking and insurance industries in Hong Kong have already adopted Open APIs, exemplified by the Open API Framework for the Insurance Sector in Hong Kong[56] and the Open API Framework for the Hong Kong Banking Sector[57].

- **OCR** is predominantly used for capturing unstructured data, converting various document formats into machine-readable data. It enables the efficient extraction of key information from self-provided documents by insurance applicants, such as names and dates of birth from identification documents, and healthcare information (e.g. handwritten notes) from medical records, including diagnoses and treatments. OCR has its limitations, particularly when dealing with documents that contain a mix of languages such as English, Traditional Chinese, and Simplified Chinese, as it can struggle to accurately recognise and process different character sets within a single document.

- **NLP/ML:** Unstructured data captured by OCR or API channels requires further processing and analysis using techniques like NLP or ML algorithms. These techniques categorise data variables based on their content or characteristics, and uncover patterns, relationships, and sentiments within the structured text. Table 15 below summarises how ML and NLP help to structure data. The approach used to convert unstructured data is determined by the specific scenario and the desired outcomes. FL trains NLP models on decentralised data, ensuring data privacy by sharing only model updates, not raw data. This approach enhances model accuracy and fosters collaborative insights, and can help achieve improved decision-making and service in industries like insurance.

Table 15 Typical ML or NLP applications for structuring data from various sources

Methods	Machine Learning (ML)/Natural Language Processing (NLP) applications
Text summarisation	Condenses large volumes of text data, such as policyholder applications, medical records, and other documents
Sentiment analysis	Analyses and understands customer feedback on and sentiment towards products/ services
Topic modelling	Categorises diverse statements into different topics (claims, policy changes, customer inquiries)
Keyword extraction	Extracts medical records, identifies key entities (company names, products, individuals), and helps detect fraudulent activities

56 An Open API framework rolled out by the IA on 18 September 2023, which enables seamless integration and data sharing between insurance companies and authorised third-party developers.

57 An Open API framework published by the HKMA on 18 July 2018, which allows Hong Kong banks to provide third-party service providers (TSPs) with access to the banking systems and retrieve specific information and services.

3.1.2 Step 2: Confidential Identity or Feature Matching

Step 2 involves identifying the same entities in data from different sources. 'Identity matching' focuses on matching personal IDs in data from different sources, while 'feature matching' matches up specific characteristics or features of an entity in data from different sources. Without this process, different data sources containing IDs and features all relating to a single entity may lead the model to treat each as relating to a separate entity. This can lead to data fragmentation, and ultimately, training failures. In the FL context, proper matching is essential for accurately consolidating data in order to make the learning process more effective and reliable.

In some FL implementations, participants use hash-based methods for identity or feature matching. However, using weak hash functions (e.g. MD5 or SHA-1) without additional protections can expose vulnerabilities that may enable reverse engineering. For instance, hashing of predictable data (e.g. names or dates) with a weak function can be cracked using rainbow tables, precomputed tables for reversing cryptographic hashes that are commonly used in password cracking. To mitigate this, a strong hash function (e.g. SHA-256) should be used and a unique salt appended to the input before hashing. Salting ensures that identical inputs produce different hashes, rendering rainbow tables ineffective unless the salt is compromised.

The choice between identity and feature matching depends on the specific scenario and the nature of the FL process. Identity matching is common in vertical federated learning (VFL), where identifiers or existing records are used to verify data from different sources. By contrast, feature matching is relevant in horizontal federated learning (HFL), where specific attributes are standardised for compatibility and consistency in model training. The automated platform advances to the model training phase once identities or relevant features are matched.

3.1.3 Step 3: Model Training and Aggregation

Model training: This is the stage at which a selected model learns from labelled training data to generate accurate predictions or decisions. In an automated system utilising FL, model training occurs on local participant servers, while model aggregation takes place on a central server where locally trained models are combined into a single global model.

The specific requirements and characteristics of a data analysis task determine whether to apply a common ML approach or a deep learning approach for model training.

(i) **ML:** Suitable for tasks with relatively straightforward data patterns and relationships. It works effectively with labelled data of a moderate size and complexity, and can analyse features to make accurate predictions. Section 3.2 offers a detailed overview of common ML processes.

(ii) **Deep learning:** Ideal for tasks with complex data patterns and relationships, such as image or text analysis. Deep learning models, particularly neural networks (NN), excel at uncovering complex patterns from large datasets, enabling highly accurate predictions.

During training, the model adjusts its internal parameters to enhance the accuracy of its predictions based on the labels of the training dataset. This process uses iterative optimisation algorithms like gradient descent to update the model's parameters[58] based on the calculated error or loss. The model continues training until it reaches a satisfactory performance level. Having learned from the training data, the model is now ready to evaluate and make predictions about new and unseen data.

Model aggregation: After local model training on the client's server, the trained parameters are shared with a central server administered by a coordinator. The coordinator's primary role is to facilitate the FL process, manage communication, and aggregate model updates. The coordinator does not access the clients' raw data or require its transmission under any circumstances, ensuring that the underlying data remains local to each client. The coordinator is typically an independent entity with no conflicts of interest, enabling decentralised training while serving as a central point of coordination, quite different from the centralised data aggregation required by conventional models. While having a coordinator might seem to contradict the decentralised nature of FL, this role is intended to support, not compromise, decentralised training. The coordinator's role includes facilitating communication, aggregating model updates and maintaining participant autonomy, thus enhancing

58 Model parameters are configuration variables that are internal to the model and learned during the training process. The specific set of model parameters depends on the type of model being used. For example, if a linear regression model is being used, the model parameters would include the slope (weight) and intercept.

the process and ensuring efficiency without centralising control. Emerging FL frameworks are continuing to explore fully distributed approaches to further minimise any reliance on central coordination.

The following steps are integrated into the coordinator's backend system and processed automatically:

i. **Collection of model parameters:** The coordinator receives the trained model parameters from the local models, which consist only of metadata such as model gradients representing the knowledge acquired by each local model.

ii. **Federated model aggregation and update:** Using FL, the coordinator performs computations on the encrypted model updates using homomorphic encryption techniques, enabling the parameters to be combined while preserving privacy and minimising the risk of exposing sensitive data. Once the model aggregation is complete, the coordinator updates the global model parameters based on the aggregated results.

iii. **Updated model distribution:** The coordinator distributes the updated global model parameters back to the local models, enabling them to learn from each other and improve their performance.

3.1.4 Step 4: Smart Decision-making

In the final step of the collaborative data analysis process, insurance companies are able to harness the trained model's predictive power to make informed decisions. By combining federated data analytics and conventional analysis models, deeper insights for various insurance tasks can be made available. This includes forecasting market trends, detecting fraudulent claims, and identifying factors that influence renewal decisions.

3.1.5 Step 5: Ongoing Assessment

Step 5 involves the ongoing assessment and evaluation of model performance in real-world environments to ensure models remain accurate, reliable, and effective over time. One effective strategy for ongoing evaluation is the champion-challenger approach, which has been widely adopted in industries like insurance and is a key component of machine learning operations (MLOps), a set of practices that automate and simplify machine learning workflows and deployments. By enabling the simultaneous evaluation of established (champion) and innovative (challenger) models, the champion-challenger approach aims to improve operational efficiency and performance, thereby facilitating the production process.

The "champion" is the current operational production model, which is performing reliably and has a proven track record. The "challenger" is a new model, often powered by advanced ML, alternative data or FL, which is tested in parallel using the same input data. If the evaluation metrics (e.g. accuracy, risk prediction, or cost efficiency) show that the challenger outperforms the champion, the challenger can replace it and become the new champion. This iterative cycle fosters continuous improvement.

The champion-challenger approach offers several advantages:

- **Integration of strengths:** Conventional models provide stability and reliability, while FL models, leveraging decentralised data, may uncover novel patterns. Combining them offers a balance of proven and cutting-edge insights.

- **Continuous improvement:** Testing FL models alongside conventional ones allows insurers to refine their predictive capabilities iteratively, and adopt superior models as they emerge.

- **Enhanced insights:** FL challengers can capture complex, distributed data patterns that conventional champions might miss, enriching analyses.

- **Risk mitigation:** Pairing innovative FL models with trusted conventional ones reduces the risks associated with using untested methodologies, grounding decisions in established strategies.

This dual approach enhances accuracy and comprehensiveness across key areas:

- **Validation & trust:** Conventional models serve as a benchmark, cross-validating results to flag anomalies and build confidence in prediction.

- **Risk management:** The reliability of conventional champions supports high-stakes decisions, while FL adapts to emerging trends.

- **Adaptation:** FL models excel at rapid adaptation to new data patterns, complementing the stability of conventional models for reporting and planning.
- **Regulatory/ethical:** Conventional models may align better with regulatory or ethical constraints, while FL evolves to meet these standards over time.

3.2 Development of Machine Learning Models

FL ensures data confidentiality and privacy by having participants train their models in a decentralised manner. The fundamental steps for training local models remain the same as in the traditional ML model training process. The following four steps summarise these processes:

Figure 11 Standard steps in a traditional ML model

3.2.1 Data Preparation and Preprocessing

Data preparation and preprocessing are the initial steps in the ML pipeline. Once data is collected, the preprocessing phase begins, which includes data integrity checks and new variable derivation.

- **Data integrity checks:** These help identify and fix data problems or errors before model training, to ensure data quality and reliability. Each attribute or field in the dataset is validated for the correct data type, and any discrepancies are flagged. The checks also identify missing values, such as empty fields or NULL values. Table 16 below outlines strategies for handling specific data integrity issues.

Table 16 Strategies for handling specific data integrity issues

Issues	Description	Strategies
Outlier	Data points that are significantly different from the rest.	• Remove if due to errors or extreme values. • Retain and manage if they represent true values.
Missing Values	Entries with no value due to input errors, equipment malfunctions, or data corruption.	• Impute using mean or median of the existing data, if the absence of value is random. • Use regression analysis to estimate based on known values, if there are correlations with other variables.
Data Anomalies and Errors	Irregularities or inaccuracies that deviate from expected patterns or behaviour.	• Employ anomaly detection, using statistical models or machine learning (ML) algorithms to identify unusual patterns. • Implement data cleaning such as deduplication, typo correction, and imputation to address data errors.

- **New variable derivations:** This involves creating new variables from existing data to enhance the model's predictive power or descriptive capability. New variables are derived through mathematical operations, transformations, or combinations of existing variables, with the aim of uncovering additional information or patterns that may not be readily apparent in the original dataset. Table 17 lists different methods for new variable derivations. By applying these methods, each party enriches its own dataset to capture local patterns. When aggregated into a global model via FL, these enhanced local models contribute to greater accuracy and nuance.

However, FL's decentralised nature presents challenges when variables depend on data held by different parties. For instance, if Party A holds data on a person's height and Party B data on their weight, direct calculation of their body mass index (BMI)—calculated as weight divided by height squared—cannot be performed without sharing raw data, a situation which FL seeks to avoid.

FL addresses this issue through specific model architecture or secure protocols. In this example, the process would work as follows:

1. **Local preprocessing:** Each party can preprocess its data individually. For instance, Party A can compute the square of the height (height^2), while Party B retains the weight.

2. **Secure aggregation:** During the FL process, these preprocessed inputs can be shared in a secure manner. The global model learns to combine these inputs without ever needing access to the raw height or weight.

3. **Collaborative insights:** By aggregating these local computations, the global model can effectively estimate BMI, despite not having direct access to the complete data from either party.

Additionally, advanced techniques like secure multi-party computation can derive such features privately, although they increase the computational cost. Table 17 lays out methods for local new feature derivations. However, FL's strength lies in its ability to generate collaborative insights from distributed data without centralising it, in a balance of privacy and utility.

Table 17 Methods for new feature derivations

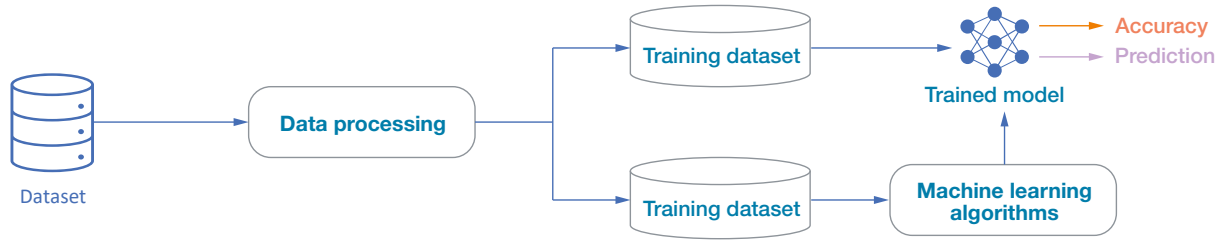
Method	Description
Mathematical Transformations	Create new or modify existing variables using mathematical functions to normalise variables or data, reduce skewness, or establish non-linear relationships. Common types include logarithmic, exponential, square root, and power transformations.
Interaction Terms	Create new variables that represent the product or combination of two or more existing variables, assessing how one predictor variable affects the relationship between another predictor and the outcome variable.
Aggregation	Combines multiple pieces of data into a single summary statistic, using methods such as summing, averaging, or finding the maximum or minimum values. This is particularly useful for time series data, where aggregation occurs over specific periods.

3.2.2 Model Training and Validation

Local machines preprocess raw data into meaningful high-quality data, which is then divided into training and testing datasets.

- **Training set:** ML algorithms construct the model and train it by feeding it labelled examples from the training set. The model is trained by learning patterns and correlations from the input data.
- **Testing set:** The testing set is used to evaluate the trained model's performance. By assessing the model's predictions against this dataset, various performance metrics, such as accuracy, can be calculated. This helps gauge how effectively the model is able to generalise to new, unseen data and how accurately it can answer specific questions or make predictions (see Figure 12 for the workflow).

Figure 12 Workflow of ML model training



It is difficult to draw a definitive conclusion about the best algorithm for use in the insurance sector because different algorithms' predictions are highly dependent on the alternative data available. Thus, selecting an appropriate ML algorithm for prediction necessitates a continual evaluation of both older and state-of-the-art ML models. Table 18 summarises and compares several typical or ML algorithms.

Table 18 Comparative analysis of ML algorithms

Algorithm	Method	Description	Pros	Cons
Logistic Regression (LR)	Probabilistic method	• Predicts the probability of a binary outcome based on one or more independent variables. • Commonly used in classification tasks with categorical outcome variables.	• Simple and easy to interpret • Efficient to train and predict • Provides probabilistic outputs	• Assumes a linear relationship between features • Less effective on data with a complex pattern • Sensitive to outliers
Naive Bayes	Probabilistic method	• Applies Bayes' theorem and assumes conditional independence among features to calculate class probabilities, selecting the most probable class as the prediction. • Quick and effective for large datasets with many features.	• Computationally efficient • Works well on high-dimension data	• Assumes independence between features which may not always be valid • Susceptible to irrelevant features

Algorithm	Method	Description	Pros	Cons
K-Nearest Neighbours (KNN)	Non-parametric method	• Uses proximity to classify or predict the grouping of a single data point. • One of the most popular and simplest classification and regression classifiers used in ML today.	• Easy to understand and implement • Works well with nonlinear decision boundaries • Can handle multi-class classification	• Sensitive to the choice of parameter value K • Requires appropriate scaling of features • Hard to explain the underlying relationships between data
Decision Tree	Tree-based method	• A flexible algorithm capable of handling classification and regression tasks by creating a tree-based model that partitions data using feature values. • Known for its interpretability and ability to handle both numerical and categorical data types, offering a versatile solution in ML for various applications.	• Easy to interpret and visualise • Robust in handling missing values and outliers	• Prone to overfitting • Sensitive to small variations
Random Forest	Ensemble learning method-Bagging	• Improves the accuracy of prediction by aggregating the predictions of multiple individual decision trees. • Robust enough to handle high-dimensional data well.	• Reduces overfitting through ensemble learning • Robust in parallel or distributed computing	• More complex than one single decision tree • Lack of interpretability compared to a single decision tree

Algorithm	Method	Description	Pros	Cons
Gradient Boosting/ XGBoost	Ensemble learning method-Boosting	• Combines weak learners (usually decision trees) to create a strong predictive model. • Through iterative training and the addition of new models, aims to correct the mistakes made by previous models, thereby improving overall prediction accuracy.	• Able to handle complex data • High predictive accuracy	• Sensitive to overfitting • Computationally expensive and time-consuming
Neural Network (NN)	Deep learning method	• A computational model inspired by the human brain, made up of interconnected artificial neurons in layers. • Excels at learning complex patterns and relationships, making it suitable for tasks like classification, regression, and image recognition. • Common types include the Multi-Layer Perceptron (MLP) for classification and regression, Convolutional Neural Networks (CNN) for image and video processing, and Recurrent Neural Networks (RNN) for sequential data tasks.	• High predictive accuracy • Easy to handle non-linear data with a large number of features	• Large network size • Computational complexity, requiring much parameter tuning • Lack of interpretability

Another approach for assessing a model's performance is cross-validation. This involves dividing the dataset into multiple folds, and using one of these folds as a validation set while carrying out training on the remaining folds. This process is repeated multiple times, using a different fold as a validation set each time. Finally, the results from each validation step are averaged to produce a robust estimate of the model's performance.

3.2.3 Model Evaluation and Prediction

For insurers, reliable assessment models should be capable of providing answers to important questions regarding the likelihood of a policyholder making a future claim, the expected severity or cost of potential claims, and the appropriate premium rate and coverage limits for a specific risk profile. Ensuring the reliability of these models requires effective techniques and practices for model evaluation. Table 19 illustrates several common metrics used to evaluate the performance of ML models, depending on the problem type.

Table 19 Overview of key performance metrics in ML

Metrics	Range[59]	Measures	Application scenarios
Area Under the Curve (AUC)	0 to 1, above 0.8 is often considered good	The area under the Receiver Operating Characteristic (ROC) curve[60]. A higher AUC value indicates better discrimination between positive and negative instances.	Ranking problems, binary classification
Accuracy	0 to 1, above 0.8 is often considered good	Proportion of correct predictions among total predictions, providing an overall measure of the model's correctness.	General classification problems where classes are balanced
Precision	0 to 1, above 0.7 is generally considered good	Proportion of true positives (TPs)[61] among all predicted positives, measuring the ability of the model to avoid false positives (FPs)[62] such as classifying legitimate emails as spam.	When false positives are costly (e.g. spam detection)
Recall	0 to 1, above 0.7 is often considered good	Proportion of TPs among all actual positives, quantifying the model's capability to avoid false negatives (FNs)[63], such as cases where patients with a disease are missed or incorrectly classified as negative.	When FNs are costly (e.g. medical diagnosis, fraud detection)
F1 score	0 to 1, above 0.7 is considered good	Combines both precision and recall, useful when there is an imbalance between the positive and negative classes.	Imbalanced classification problems
MSE	Always non-negative	The average of the squares of the errors between actual values and predicted values, making it easier to compute the gradient.	Regression tasks
KS index	0 to 1, a higher value indicates better performance	Measures the maximum difference between the positive and negative classes	Binary classification problems

59 The threshold for "good" depends on the specific problem being addressed.
60 The ROC curve is a graphical plot that showcases how well the model performs. It visualises the relationship between the true positive rate (TPR) and the false positive rate (FPR) over all possible acceptance thresholds.
61 True positives (TPs) occur when the model's prediction matches the truth.
62 False positives (FPs) are cases where the model incorrectly predicts a positive outcome for negative instances.
63 False negatives (FNs) occur when the model predicts a negative outcome, while the true class is positive.

Once an ML model has been trained and evaluated, it is ready to make predictions on new, unseen data. Model prediction is the process of using the trained model to generate output or make decisions based on input data.

3.2.4 Model Explainability

The pursuit of improved predictive accuracy has led to increased model complexity, with deep learning being at the heart of many state-of-the-art ML systems. However, this complexity comes at a cost: deep learning models are often opaque and difficult to interpret, earning them the nickname "black boxes".

Lack of interpretability can be a major obstacle to trust, particularly in sensitive areas like finance, transportation, and healthcare. For instance, a bank using AI to make credit decisions must provide clear reasons for loan denials, or customers may lose trust in the system and feel unfairly treated. Similarly, autonomous vehicles must be able to explain their driving decisions, as a lack of explainability can lead to scepticism and a reluctance to adopt the technology. In healthcare, if doctors and patients cannot understand the reasons for the recommendations generated by ML algorithms, they are likely not to trust these recommendations. The need for trustworthy, fair, robust, and high-performing models has led to the rise of Explainable Artificial Intelligence (XAI). XAI aims to make AI models transparent, interpretable, fair, and robust by providing insights into their decision-making processes, addressing biases, ensuring reliability, and fostering user trust.

Several tools and libraries have also emerged that are helping make AI models more transparent and interpretable. Examples include:

- Python software libraries such as Scikit-learn which can analyse AI models and explain which factors or features in the data are most important for the model's predictions. This helps users understand how the AI is making decisions.
- Tools such as Explain Like I'm 5 which can break down the reasoning behind an AI model's output in plain language and explain it in simple terms. This makes the inner workings of complex AI models more accessible.
- Software that uses mathematical concepts like "Shapley values" to visualise the influence of different input factors on an AI model's output, offering insights at both local and global levels. This provides a more comprehensible way to interpret how the model is arriving at its conclusions.
- Libraries such as Local Interpretable Model-Agnostic Explanations can explain the predictions of any type of AI model, even ones that are very complex and extremely difficult to understand. This increases transparency by providing detailed insights into individual predictions.

These types of interpretability tools are making it easier for humans to understand and trust the reasoning behind AI-powered decisions and predictions, thus making AI models more transparent and accountable.

3.3 Privacy-enhancing Techniques for Federated Learning

FL is a prominent privacy-enhancing technique, but it is not entirely immune to the risks of data or model leakage under adversarial attacks, as discussed in Part Two. The real-world environment often presents complex challenges in this respect. Despite this, the Hong Kong government has recognised the importance of promoting the digital economy and building a robust data trading ecosystem within the region, making further enhancements to the reliability and practicality of FL a crucial focus area.

This section discusses privacy-enhancing techniques (PETs) for FL systems and explores specific PETs that address the challenges associated with identity matching in FL environments.

3.3.1 Key Privacy-enhancing Techniques

3.3.1.1 Overview of PETs

Common examples of mainstream PETs include secure multi-party computation (SMPC), differential privacy (DP), homomorphic encryption (HE), and confidential computing. Table 20 summarises each type and discusses their pros and cons.

- **SMPC**

SMPC enables collaborative data analysis while preserving privacy by allowing separate parties to jointly derive insights without revealing specific data values. One key technique is secret sharing, which involves dividing sensitive data into shares that are distributed among multiple parties, ensuring that no single party has complete access to the original data.

In the context of the insurance industry, SMPC allows insurance companies and data providers to perform joint computations on encrypted data, facilitating collaborative model training and analysis without exposing sensitive information. This approach enables the encrypted results to be shared, helping insurance companies benefit from external insights while complying with data privacy and regulatory requirements.

- **DP**

Compared to data anonymisation, DP provides a quantifiable privacy guarantee by introducing random noise into the dataset. This is measured by a parameter called epsilon. A smaller epsilon value indicates a stronger privacy guarantee, as it represents more noise and a lower risk of re-identification. In FL, each device or organisation applies DP techniques to add controlled noise to their local data before sharing it with a central server. This enables insurance companies to protect policyholder identities and protect customer privacy, while still enabling collaborative data analysis.

- **HE**

HE is a cryptographic technique that allows computations to be performed directly on encrypted data, without the need for decryption. Unlike SMPC, which involves collaboration among multiple parties, HE integrates encryption and decryption processes into the computation itself. This minimises the need for additional communication and interaction between parties during computation, enhancing efficiency.

- **Confidential computing**

Confidential computing enables data to be processed within a secure environment, enhancing data security by preventing unauthorised access during computation. It employs two key security techniques: isolation, which protects sensitive information while in use, and remote attestation, which verifies this protection and the data's intended purpose before computation begins.

Trusted execution environments (TEE) are essential for confidential computing, as they provide for the hardware-enforced isolation of sensitive code and data. TEEs are secure enclaves that protect against external tampering, including tampering by privileged system software such as an operating system or hypervisor, while maintaining confidentiality and integrity during execution.

Table 20 Summary of key PETs and their pros and cons

PET	Mechanism	Pros	Cons
Secure multi-party computation (SMPC)	Protects parameters sent from participants to ensure that they do not reveal their inputs.	• Privacy preservation • Data security • Decentralised computation	• Communication overhead • Scalability limitations • Highly rely on practical factors like network bandwidth
Homomorphic encryption (HE)	Encrypt local parameters from all participants. The coordinator server receives an encrypted global model which can only be decrypted if enough local models have been aggregated.	• Data privacy • Secure computing • Flexible applications	• Computational overhead • Complex key management • Lack of widely accepted standards and protocols
Secure communication protocols	Protocols between clients and between clients and the central server. Utilises protocols to prevent man-in-the-middle attacks, eavesdropping, and tampering.	• Data confidentiality • Ensuring the information integrity • Standardised communication protocols	• Complex to implement • Performance overheads • Vulnerability to attacks
Differential privacy (DP)	Adds noise to a particular individual's data to hide the fact that the individual's data was used in the training task.	• Privacy guarantee • Resilience to data breaches • Algorithmic foundations	• Accuracy trade-off • Limited application scenarios • Difficulty in parameter tuning
Confidential computing	Uses trusted execution environments (TEE) to isolate and secure the execution of code and data.	• Enhanced security	• Complex implementation • Limited availability • Performance overheads

3.3.1.2 Fast-Training strategy module (FSTM)

Privacy-preserving techniques often require extra computations, such as noise addition, data transformations, cryptographic proofs, and secure communication protocols. These extra computations introduce additional computational overheads, as a trade-off for protecting privacy. Operations such as encryption and decryption involve complex mathematical calculations such as modular exponentiation and multiplication, while the generation of cryptographic keys, including public-private key pairs, is computationally intensive.

The FTSM[64] is proposed as a way of making privacy-enhancing techniques more practical for a real-world FL platform. This is a module that aims to improve the efficiency and computational performance of the FL training process by enabling faster training while preserving privacy guarantees. Table 21 provides a comparison of the characteristics of the FSTM and mainstream PETs.

Table 21 Comparison of the FTSM and mainstream PETs

Aspect	Homomorphic encryption (HE)	Differential privacy (DP)	Fast-training strategy module (FSTM)
What is Generated?	Public-private key pairs for encryption	Noise, such as Laplace noise or Gaussian noise	Beaver triple matrix for training
Key Computation Feature	Performs computation on encrypted data without decrypting data	Adds noise to the computation parameters or output results	Performs computation on a matrix with randomness
Type of Delivered Data	Encrypted model parameters, such as gradient	Computation parameters with added noise	Computation matrix manipulated by the random matrix
Characteristics	High computational complexity Large ciphertext size	Noise might impact model accuracy and performance	Matrix speeds up the calculation process but increases the implementation complexity Comprehensive design is required before implementation

As shown in the table, HE algorithms can suffer from performance inefficiencies due to the overheads associated with maintaining public-private key pairs. DP, which addresses privacy concerns by introducing noise into calculation results, can compromise accuracy and performance.

By comparison, the FTSM's use of matrix calculations delivers a significant improvement in computational speed, mainly due to the parallelisability of matrices and their inherent mathematical properties. Leveraging these properties allows for efficient parallel computations, resulting in faster execution times and enhanced computational efficiency.

64 Yang Liu et al., A Secure Federated Transfer Learning Framework, IEEE Intelligent Systems, 2020.

3.3.2 Secure Identity Matching for Utilisation of Alternative Data

Once each participant has uploaded their data to their respective local servers, conducting secure data matching (e.g. identity matching or feature matching) while ensuring the protection of sensitive information becomes a crucial challenge. Traditional identity matching methods often involve sharing personally identifiable information (PII), posing significant risks to data privacy and security.

Privacy-preserving identity matching techniques have been developed to address these concerns. One is private set intersection (PSI), which enables parties to find common elements or matches between their datasets without revealing any non-matching data, allowing for secure collaboration and identity reconciliation without directly exposing sensitive information.

3.3.2.1 Overview of PSI

- **PSI:** PSI is a SMPC that allows organisations to identify common elements in their datasets without revealing the specific contents of their respective datasets. PSI only shows the shared features across different datasets, facilitating the linking of individuals or data elements across organisations for various use cases. It reduces privacy risks by only revealing the standard features shared between two datasets, without requiring both parties to disclose their entire datasets to each other.

- **Classification of PSI protocols:** Achieving the secure intersection of two sets without compromising the confidentiality of any information except the resulting intersection is a significant challenge for secure computation. Several techniques have been suggested to address this problem, including the efficient yet insecure naïve hashing solution, protocols that rely on a partially trusted third party, protocols based on public key, circuit-based PSI, and Oblivious-Transfer PSI. Each category is discussed in detail below (see Table 22[65]):

Table 22 Classification and details of PSI protocols

PSI protocols	Key approach	Pros	Cons
Naïve hashing	Uses a hash function to hash its stored elements	• Efficient in run time and communication	• Vulnerable to a brute-force attack • Hash collisions
Server-aided/ Third party-based	Employs a third party to achieve better performance. The server could be semi-honest, covert, or malicious.	• Reduced communication overheads	• Secure only if the third party does not collude with any of the other clients
Public key cryptography-based	Encrypts the elements using public-key cryptography, such as the Diffie-Hellmann key agreement and blind RSA	• Does not require a trusted third party or central server	• Requires proper key management • Large communication overheads
Generic protocol/ Circuit-based	Uses generic secure computation	• High security	• Circuit-based approach requires expensive computation and communication
Oblivious transfer-based	The receiver obtains one out of multiple potential messages from the sender without the sender learning which specific message was chosen or revealed to the receiver.	• Efficient communication	• Needs intensive computation • Requires large computational resources

65 Benny Pinkas et al., Scalable Private Set intersection Based on OT Extension. ACM Transactions on Privacy and Security, 2018.

- **Basic PSI process:** In general, there is no single universally accepted standard process for PSI. The following steps provide a simplified overview of how PSI typically works:

1. Setup: The participating parties agree on a cryptographic scheme and establish their private sets.

2. Encryption: Each party encrypts its set using a secure encryption scheme that allows for set membership testing without disclosing the actual elements.

3. Exchange: The encrypted sets are exchanged between the parties, ensuring the preservation of privacy. This exchange can occur directly between the parties, or through a trusted intermediary.

4. Intersection: Each party performs operations on the received encrypted sets to determine the common elements. This typically involves comparing the encrypted values and identifying matches.

5. Decryption: Once the intersection has been identified, the parties can decrypt the matching elements to reveal the details in their sets.

- **PSI applications:** PSI has a wide range of real-world applications across various domains. For instance, PSI can be used in fraud detection and anti-money laundering, online recommendation systems, confidential data sharing, border protection and no-fly lists, network security operations, customer list intersections for marketing, medical research and patient data analysis, multi-party access control, enterprise network auditing, and many more situations. Some practical examples are listed below:

    - **Insurance companies:** Multiple insurance companies can use PSI to find intersections in their customer lists, enabling them to identify shared customers without disclosing individual customer information.

    - **Healthcare providers:** PSI can facilitate the secure exchange of medical information between healthcare providers in a way that complies with privacy regulations. It ensures that sensitive patient data remains protected while enabling efficient data sharing for improved healthcare outcomes.

    - **Social network applications:** In a social network application, two users can use PSI to discover common friends without revealing other friends that are not in the intersection. This preserves privacy while still enabling social connections to be established.

However, PSI can introduce certain risks, such as risks of re-identification due to inappropriate intersection sizes or over-analysis. In our proposed framework, a CIMM has been designed for identity matching purpose to mitigate such risks. The CIMM uses secure hashing combined with random number generation to protect identities by transforming them into hashed representations, with only partial intermediate results sent to a neutral coordinator who has no access to the raw data. The coordinator uses HE to compute on encrypted data, enabling identity matching without revealing sensitive information.

3.3.2.2 Confidential identity matching module (CIMM)

• Problems with general FL frameworks

Lack of a third party acting as coordinator

In a VFL framework that lacks a third party acting as a coordinator, the hashed values of raw identities are exchanged between parties A and B, or party A sends the hashed values to party B to compare the matches. However, this approach poses risks to the privacy and security of the data because the partner party now holds the hashed values, leaving them potentially vulnerable to attackers. For instance, an attacker may employ brute-force attack techniques to uncover the original data.

Untrusted third party

A third party is a critical component of the FL framework. However, if the third party is not trustworthy, it could collude with one of the parties, leading to biased models, data breaches, or other security threats. For example, the third party could intentionally introduce bias into the model, or leak sensitive data to a competitor.

Reverse engineering with a third party

Even with a third party, the encrypted data received from parties A and B can still be vulnerable to reverse engineering attacks. Reverse engineering is the process of analysing the encrypted data to uncover its hidden patterns and relationships, compromising the privacy of the data and potentially leading to data breaches. For instance, an attacker could use clustering algorithms to identify patterns in the hashed data and infer sensitive information about the data owners.

• CIMM

The focus of our CIMM is identity matching, which involves comparing and correlating data from various sources while maintaining privacy. It employs a hash function and the HE technique to securely match identities across different clients, as well as a neutral third party to distribute the matched results, either Boolean or operator, back to the clients. Some highlight features are as follows:

1. **Secure hashing with random number generation:** The CIMM employs hashing techniques to transform the original identities into hashed representations. It also includes random number generation to enhance the privacy and security of the identities. This combination provides an additional layer of security for sensitive information, with only a portion of the intermediate computation results sent to the coordinator.

2. **Neutral coordinator:** The coordinator computes the difference between the clients' encrypted identities and returns the result. This process enables clients to determine matches or non-matches without revealing sensitive information. In this research, the coordinator (the Insurance Authority) had no access to the raw data stored on the clients' devices. Further details about the coordinator's role can be found in Section 3.1.3.

3. **HE for secure computations:** The coordinator performs computations directly on the encrypted data using HE techniques without the need for decryption.

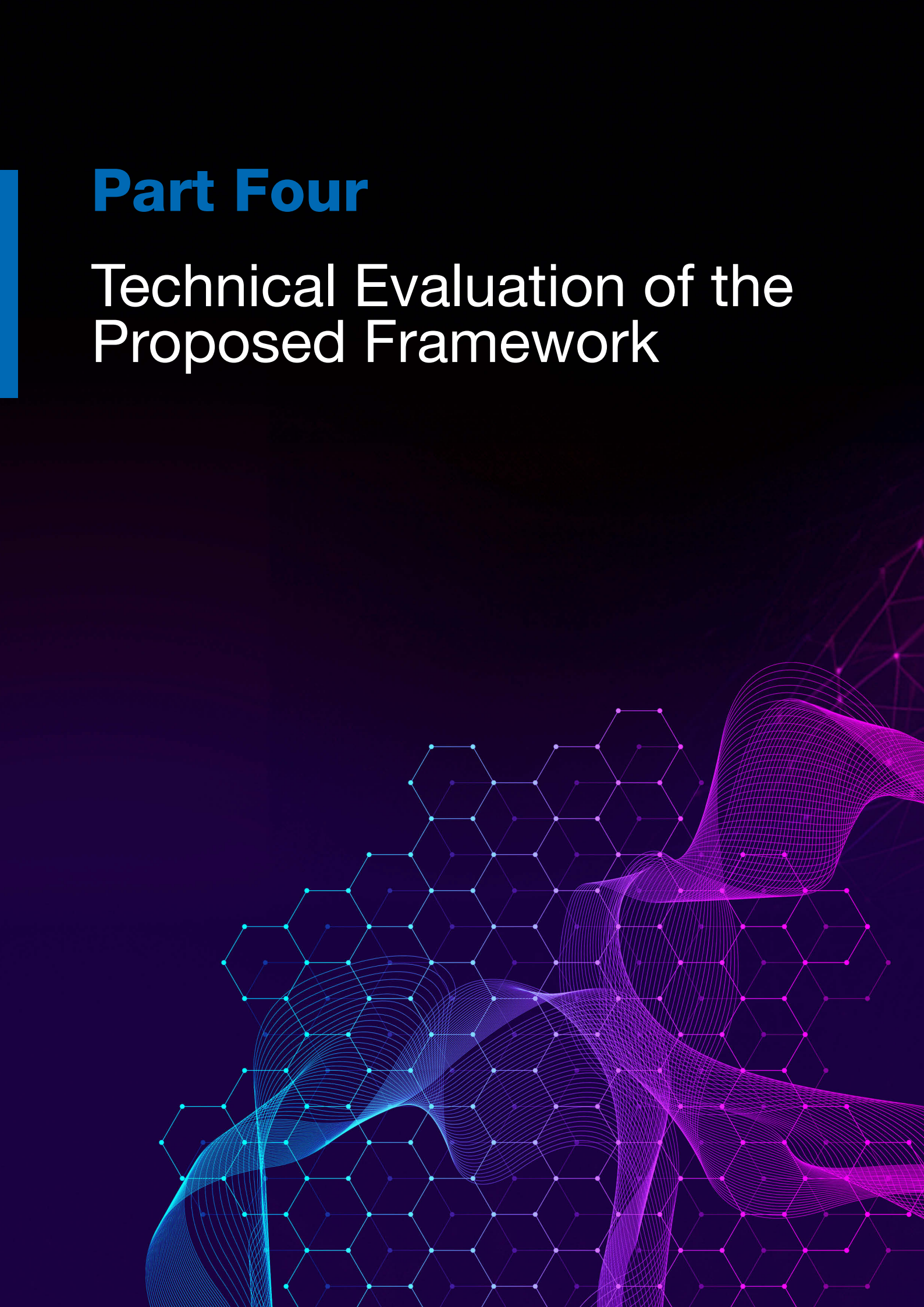
Table 23 provides a detailed overview of how the CIMM addresses the identity-matching problems described above. It highlights the role of the neutral coordinator, the use of encrypted data exchange, and the protection it offers against reverse engineering.

Table 23 How the CIMM solves key problems

Problems	How our module solves the problems
Lack of a third party acting as coordinator	The CIMM incorporates a coordinator to oversee the matching process. Hashed identities are encrypted, adding an extra layer of protection to safeguard the privacy of the data.
Untrusted third party	The coordinator's neutrality ensures that it cannot collude with any of the clients. Even if collusion were to occur, the coordinator does not have access to the original identities.
Reverse engineering with a third party	The coordinator receives the differences between the encrypted identities rather than the encrypted identities themselves.

Part Four

Technical Evaluation of the Proposed Framework



Part Four:
Technical Evaluation of the Proposed Framework

This part evaluates the technical feasibility of using the FL framework introduced in Part Three in the insurance industry. The objective is to assess the framework's performance and efficacy using real-world insurance-related datasets.

Before the PoC work was undertaken, different ML algorithms were first evaluated on a set of open-source datasets. These were grouped into three categories based on their source characteristics, namely generic policyholder information, premium data, and finance data. The experiments were conducted to evaluate how alternative data sources and varying data volumes influence different ML models' performance. Models were trained and tested with different data groups, and the model performance was assessed with the Area Under the Curve (AUC) metric. An assessment of the fast-training strategy module (FTSM) was also undertaken.

4.1 Introduction to the Experiments

The experiments tested eight ML algorithms: Logistic Regression (LR), Naïve Bayes, K-Nearest Neighbours (KNN), Decision Tree, Random Forest, Gradient Boosting, XGBoost, and Neural Network (NN). Each algorithm was independently trained with hyperparameter tuning to optimise configurations. Detailed algorithm descriptions are in Section 2 of Part Three, while tuning specifics are omitted for brevity.

Conducted in a Jupyter Notebook, the experiments involved splitting the data into parts that acted as diverse data sources from different parties. Performance outputs were exported for analysis and their effectiveness was assessed using ROC curves and the AUC, which serves as a reliable measure of classification performance.

The aim was to assess the likelihood of a policyholder paying the 13th-month premium at the new business stage. Accurate predictions of policy renewals are important for insurers, helping them design better policies and improve their customer retention strategies. The experiment consisted of two parts:

- **Experiment one:** Evaluated how different algorithms handled alternative data, specifically in a vertical federated learning (VFL) scenario, by dividing the dataset into groups based on generic, premium and finance characteristics, and using AUC for performance assessment.

- **Experiment two:** Assessed the effectiveness of enhancing data volume in model training within a horizontal federated learning (HFL) scenario, maintaining AUC as the consistent evaluation metric.

4.2 Data Overview

This experiment employed an insurance dataset obtained from Kaggle[66], consisting of 100,000 records, 38 features, and one binary label indicating whether the 13th-month premium was paid or not. To ensure the data quality, 28 important features and 57,580 rows of records without missing values were selected. This cleaned data was categorised into three distinct groups based on the data properties. Table 24 provides a breakdown of this categorisation from VAR1 to VAR28, presenting the division of the raw dataset into three categories, namely generic data (containing application life-assured basic information, agent information and health-related data), premium information data, and financial information data.

To further refine the input data for modelling, standardised operations were applied on certain features, such as Applicant's Policy Annualised Premium, Applicant's Policy Sum Assured, and Application Life Assured Income. This step was necessary because these features initially had varying scales. For example, the Policy Sum Assured had a wide range of values, from 0 to 700,000,000. Standardisation rescaled these features to a more suitable range for effective model training.

66 Benny Pinkas et al., Scalable Private Set intersection Based on OT Extension. ACM Transactions on Privacy and Security, 2018.

Additionally, the dataset contained some categorical variables that cannot be directly processed by most ML algorithms. Specifically, there were 11 categorical variables in the generic data category, 7 in the premium data category, 2 in the finance data category, and 1 categorical target variable, making a total of 21 categorical variables. For instance, the Policy Product Category, which includes categories such as Pension, Protection, Savings, Investment, and Children's Plan, was transformed into numerical values. This conversion enabled the ML models to interpret the categorical data and incorporate it into their learning process. To convert these categorical variables into a numerical format, the use of encoding techniques was necessary.

These steps of standardising features and encoding categorical variables were crucial for refining the input data, ensuring that the models could effectively learn from the data, capture meaningful patterns, and make accurate predictions.

Table 24 Variables of the insurance dataset for predicting 13th-month payment behaviour at the new business stage

Variable flag	Category name	Variable detail	Variable Type
VAR0	Identification	Masked Policy Identifier	Unique ID
VAR1	Generic data (Including application life assured basic information, agent information and health-related information)	Application Life Assured Age	Numerical
VAR2		Application Life Assured Education	Categorical
VAR3		Application Life Assured Gender	Categorical
VAR4		Application Life Assured Industry	Numerical
VAR5		Application Life Assured Marital Status	Categorical
VAR6		Application Life Assured Nationality	Categorical
VAR7		Application Life Assured Occupation	Categorical
VAR8		Application Life Assured Residential Status	Categorical
VAR9		Application Life Assured State	Categorical
VAR10		Application Life Assured City	Numerical
VAR11		Application Life Assured City Tier	Numerical
VAR12		Mapped Agent Branch	Categorical
VAR13		Mapped Agent Vintage	Categorical
VAR14		Life Assured Alcohol Declaration	Categorical
VAR15		Life Assured BMI	Numerical
VAR16		Life Assured Smoker Declaration	Categorical
VAR17	Premium data	Applicant's Policy Rider Opted Flag (Y/N)	Categorical
VAR18		Application's Payment Frequency*	Categorical
VAR19		Applicant's Policy Annualised Premium	Numerical
VAR20		Application Sourcing Channel	Categorical
VAR21		Application's Policy Contract Branch	Categorical
VAR22		First Premium Payment Type	Categorical
VAR23		Applicant's Policy Product Category	Categorical
VAR24		Policy Product Name	Categorical
VAR25		Policy Sum Assured	Numerical
VAR26	Finance data	Application Life Assured Income	Numerical
VAR27		Application's Policy Price Sensitivity (Y/N)	Categorical
VAR28		Auto Debit of Premium Opted Flag (Y/N)	Categorical
TARGET	/	Paid the 13th-month Premium at New Business Stage (Y/N)	Categorical

*: Application's Policy Premium Payment Frequency
/: not applicable

4.3 Experiment One: Impact of Alternative Data

To assess how well the model performed with alternative data, including generic, premium-related, and finance-related data, the three categories in Table 24 were combined into distinct groups to examine the impact on the likelihood of policyholders paying the 13th-month premium, as presented in Figure 13.

The experiment used Group 1 data, made up of generic data only, to examine the relationship between insurance traditional data and the likelihood of policyholders paying the 13th-month premium. Building upon Group 1, Group 2 expands the data complexity by incorporating the premium data to determine whether such data has an impact on the target variable. Group 3 represents the highest level of data complexity, combining generic data, premium data, and finance data. The experiments for evaluating the impact of the three types of alternative data are illustrated in Figure 13.

Figure 13
Description of data Groups 1-3 for VFL

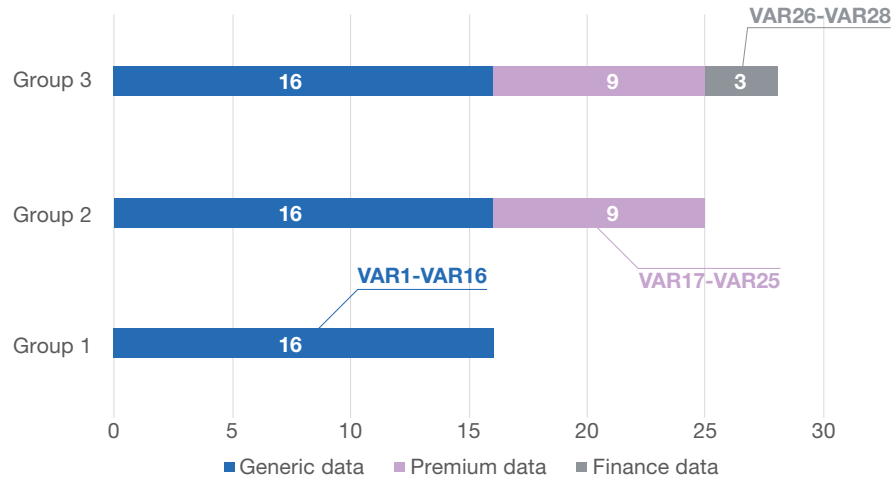
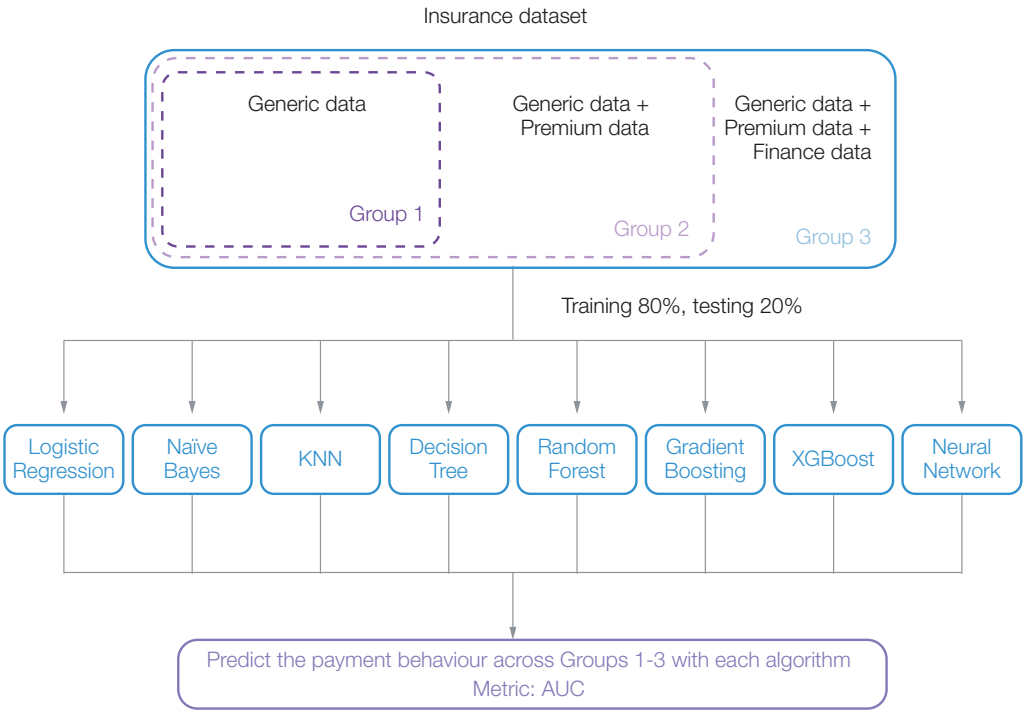


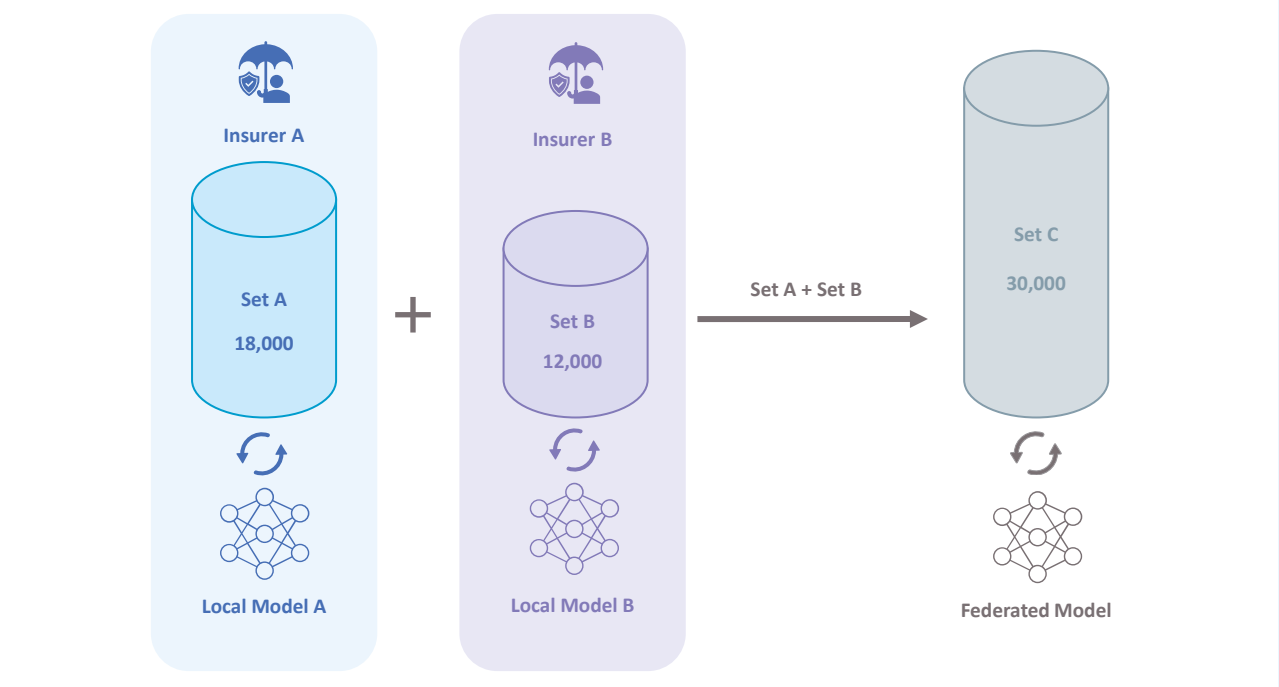
Figure 14 Experiments designed to evaluate the impact of alternative data across Groups 1-3



4.4 Experiment Two: Impact of Data Volume

A scarcity of comprehensive data may hinder insurance providers from thoroughly studying their customers' behaviour patterns. This challenge can be effectively addressed through data collaboration among insurance companies and the application of HFL. In HFL, participants share the same feature space, using a common set of features such as policy type and premium amount. Although the actual data samples differ among insurers, each insurer has data on these common features.

Figure 15 Description of Set A, Set B, and Set C for HFL



To investigate the impact of data volume on model training, this experiment split the insurance dataset into two sets. This represented a two-collaborator federation, in which Insurer A holds 18,000 samples (Set A) and Insurer B holds 12,000 samples (Set B). Through FL, Insurers A and B could collaboratively develop a global model by aggregating their datasets (Set C, 30,000 samples). The performances of a range of ML models on Set A, Set B, and Set C were evaluated separately to compare the model performances of local models and the federated model.

4.5 Evaluation Results

4.5.1 Performance (AUC scores) Using Different Machine Learning Algorithms

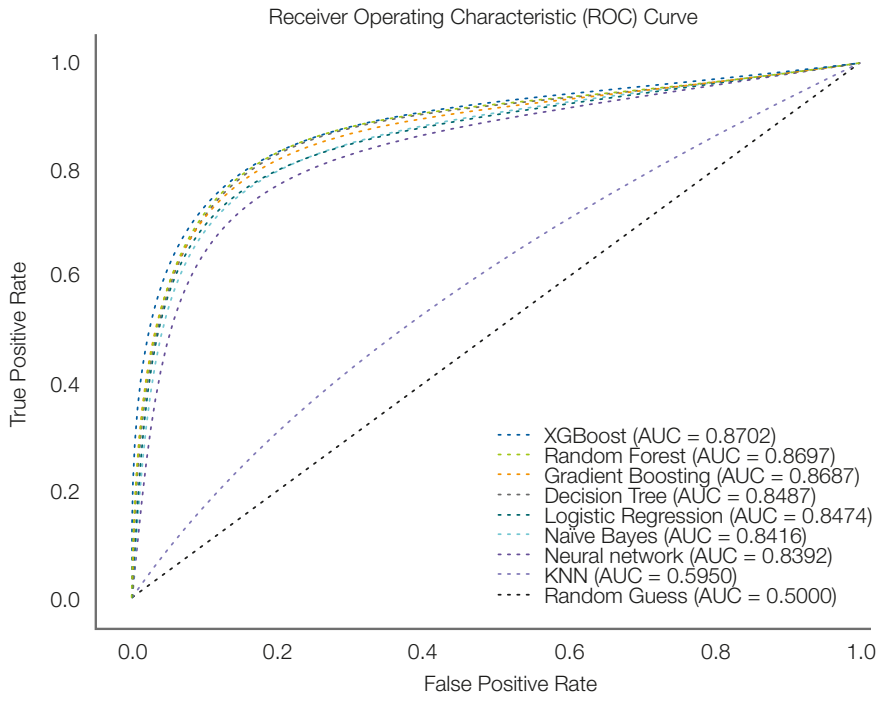
Experiment One focused on assessing the effectiveness of applying alternative data to the insurance business. The AUC metric was applied to evaluate the performance of the models. Generally, an AUC score above 0.7 is considered acceptable.

- **AUC scores of different ML algorithms on the single data group**

Figure 16 illustrates the ROC curves depicting the predictions of each ML algorithm of policyholders paying or not paying the 13th-month premium on data Group 3 (full dataset). Generally, an AUC of 0.5 indicates a model that performs no better than random guessing. An AUC of 1 implies that the model has perfect discrimination ability, being considered an ideal classifier. There is no specific threshold for the AUC that indicates a well-functioning model.

In general, modern model mechanisms – such as ensemble learning algorithms like Random Forest, XGBoost, and Gradient Boosting, demonstrated higher predictive power than traditional models like KNN and Naïve Bayes. However, these latter traditional methods have been used for a long time and have solid theoretical foundations. Traditional classification mechanisms like Decision Tree and LR also performed reasonably well on this task.

Figure 16
ROC curves of ML algorithms applied to data Group 3 (full dataset)



- **AUC scores of ML algorithms across the different data groups**

Table 25 presents the AUC scores for various ML algorithms in predicting the likelihood of policyholders paying the 13th-month premium at the new business stage across data Groups 1, Group 2, and Group 3. KNN achieved the lowest AUC scores, and XGBoost the highest (highlighted).

Table 25 AUC scores of ML algorithms across data Groups 1-3

Algorithm \ Group	Group 1 Generic data	Group 2 Generic data + Premium data	Group 3 Generic data + Premium data + Finance data
Logistic Regression (LR)	0.6079	0.6439	0.8474
Naïve Bayes	0.5825	0.6265	0.8416
K Nearest Neighbours (KNN)	0.5745	0.5817	0.5950
Decision Tree	0.6072	0.6415	0.8487
Random Forest	0.6469	0.6951	0.8697
Gradient Boosting	0.6566	0.7003	0.8687
XGBoost	0.6572	0.7034	0.8702
Neural Network (NN)	0.5923	0.6241	0.8392

a. Improved predictive performance with increased features and variables

AUC values generally increase from Group 1 to Group 3 across all models. The Group 1 dataset, with the least information, yielded the lowest AUC scores, while the addition of premium data (Group 2) resulted in slightly higher scores.

b. XGBoost outperformed other algorithms across the data groups

XGBoost consistently achieved the best performance of all the algorithms across all three data groups. XGBoost's Group 3 score was 0.8702, indicating good predictive performance.

c. KNN performed poorly in high dimensions and showed limited improvement with additional data

The KNN algorithm performed the worst of all the models, primarily due to its sensitivity to "the curse of dimensionality", where performance degrades as the feature count increases. Compared to other algorithms, KNN showed limited improvement with additional data. When there are many features, data points spread out, making it harder to find "nearest neighbours." Furthermore, KNN relies heavily on feature similarity, so if new features cannot be meaningfully discriminated amongst, its performance may not improve significantly.

d. NN outperformed KNN but lagged behind LR due to training variability

NN outperformed KNN but was surpassed by LR. NN's performance variations can be attributed to several sources of randomness during training, including parameter initialisation, sample selection, and neuron dropout, as well as the characteristics of the loss function.

e. Finance data significantly boosted model performance, except for KNN

This dataset, which includes generic, premium, and financial information, achieved the highest AUC scores among the three groups. The finance variables captured essential patterns and insights that were missing in the generic and premium features, enabling the models to better understand underlying relationships and make more accurate predictions.

- **Conclusions**

Overall, the different ML algorithms exhibited varied levels of predictive power. The progressive improvement in AUC scores across Groups 1-3 highlights the value of utilising diverse data sources, such as finance and premium information, to build more accurate and robust predictive models for the insurance industry.

4.5.2 Feature Importance of ML Algorithms

To identify the features that contributed most to the ML models used in the experiment, Shapley Additive Explanations (SHAP) were applied to extract feature importance. This powerful technique interprets the output of any ML model, regardless of its architecture or learning algorithms. However, calculating SHAP values in FL can be computationally expensive, especially for models with a large number of features, as it involves evaluating all possible combinations of the features.

Table 26 presents the top 10 most important features for the ML algorithms, excluding KNN, NN, and Gradient Boosting. Feature importance is not relevant for KNN and NN, as these rely on different mechanisms and do not provide explicit importance measures. Gradient Boosting is excluded because the top features identified by XGBoost are likely to be similar in a Gradient Boosting model. This is because both algorithms iteratively train weak models to build a strong predictive model, albeit with different regularisation techniques.

Table 26 Top 10 most important features in the algorithms

Rank	Logistic Regression	Naïve Bayes	Decision Tree	Random Forest	XGBoost
1	VAR28	VAR28	VAR28	VAR28	VAR28
2	VAR17	VAR27	VAR26	VAR25	VAR17
3	VAR13	VAR18	VAR25	VAR17	VAR20
4	VAR18	VAR17	VAR18	VAR26	VAR12
5	VAR27	VAR11	VAR20	VAR13	VAR25
6	VAR25	VAR25	VAR10	VAR20	VAR18
7	VAR11	VAR2	VAR13	VAR10	VAR22
8	VAR20	VAR3	VAR7	VAR18	VAR26
9	VAR23	VAR7	VAR23	VAR12	VAR10
10	VAR12	VAR26	VAR24	VAR7	VAR13

a. **Feature importance variation:** The importance of specific features differs among algorithms. For instance, VAR25 (Policy Sum Assured) is the second most important feature in Random Forest but is less significant in other algorithms.

b. **Consistent importance of VAR28:** VAR28 (Auto Debit of Premium Opted Flag) was consistently ranked as one of the most important features across all algorithms due to its strong influence on the likelihood of policyholders paying the 13th-month premium.

c. **Lack of health-related features:** None of the important features in Table 26 are health-related, such as VAR14 (Life Assured Alcohol Declaration), VAR15 (Life Assured BMI), and VAR16 (Life Assured Smoker Declaration) in the generic data, suggesting that health information may have less impact on the likelihood of policyholders paying the 13th-month premium. However, this could change if the focus is shifted to assessing risk profiles.

4.5.3 Evaluating the Impact of Data Volume

This section presents an analysis of the impact of increased data volume on the performance of various ML algorithms.

Table 27 AUC scores for the impact of data volume on ML algorithms

Model Set	Logistic Regression	Naïve Bayes	KNN	Decision Tree	Gradient Boosting	Random Forest	XGBoost	Neural Network
Set A	0.8551	0.8519	0.5569	0.8454	0.8676	0.8746	0.8704	0.8233
Set B	0.8542	0.8472	0.5706	0.8383	0.8671	0.8696	0.8713	0.5343
Set C	0.8555	0.8498	0.5623	0.8507	0.8760	0.8768	0.8746	0.8399

a. **AUC findings:** Table 27 shows that most algorithms achieved higher AUC scores with the combined dataset (Set C) compared to Set A and Set B, indicating that increased data volume can enhance model performance. Combining heterogeneous data sources helps overcome the performance limitations of isolated datasets.

b. **Set C performance:** Set C, which includes Set A and Set B, consistently generated the best performance by all algorithms, except for Naïve Bays and KNN. Notably, Naïve Bayes in Set C had a lower AUC score than Set A, suggesting that simply adding records does not always improve model performance. This decline is due to the algorithm's strong reliance on the assumption of feature independence, which can be compromised when more samples are added.

c. **KNN performance issues:** As discussed in the preceding section, the KNN model's sensitivity to "the curse of dimensionality" is the reason for its performance being near to random (since its AUC score is close to 0.5). Misleading information, imbalanced data, and duplicated records are among the factors that can introduce noise and negatively impact accuracy and the AUC.

d. **NN limitations:** NNs excel in handling unstructured or complex datasets, such as those relating to image recognition, computer vision, NLP, and time series forecasting, where traditional algorithms may struggle to find patterns or make accurate predictions. However, in this experiment, despite their strong feature extraction capabilities, NN underperformed compared to traditional algorithms (showing a near random result of 0.5343 on Set B), probably because the task was not complex enough to fully utilise its strengths.

• Conclusions

Most algorithms demonstrated acceptable performance but showed varying predictive capabilities. While FL is effective in many scenarios, it may have limitations within certain parts of the insurance value chain, and adding additional data sources does not always enhance model performance.

The selected algorithms showed diverse predictive abilities, emphasising the importance of selecting the right algorithm based on data characteristics and the specific problems being addressed. LR is ideal for insurers facing data scarcity, due to its simplicity, interpretability, and faster results compared to NNs, making it more efficient in certain scenarios.

4.5.4 Evaluation of the Fast-Training Strategy Module (FTSM)

Employing encryption in FL can significantly increase the processing time. To address this issue, an FTSM is proposed to accelerate training while maintaining privacy, as detailed in Section 3.3.1.2. The efficiency and effectiveness of the FTSM was evaluated using open-source data. A comparison was made between the training time per iteration of the existing approach using HE and the FTSM, with LR serving as the benchmark. The results, shown in Table 28, indicate that the FTSM outperforms HE in three key areas:

- **Training time:** The module significantly reduced the training time per iteration, achieving speeds from approximately 3.5 to 17.6 times faster than the existing approach.
- **Efficiency:** The module demonstrated superior performance in terms of speed and efficiency, leading to potential cost savings and increased productivity in data processing tasks.
- **Versatility:** The module is applicable to different datasets, offering a versatile solution for data processing tasks.

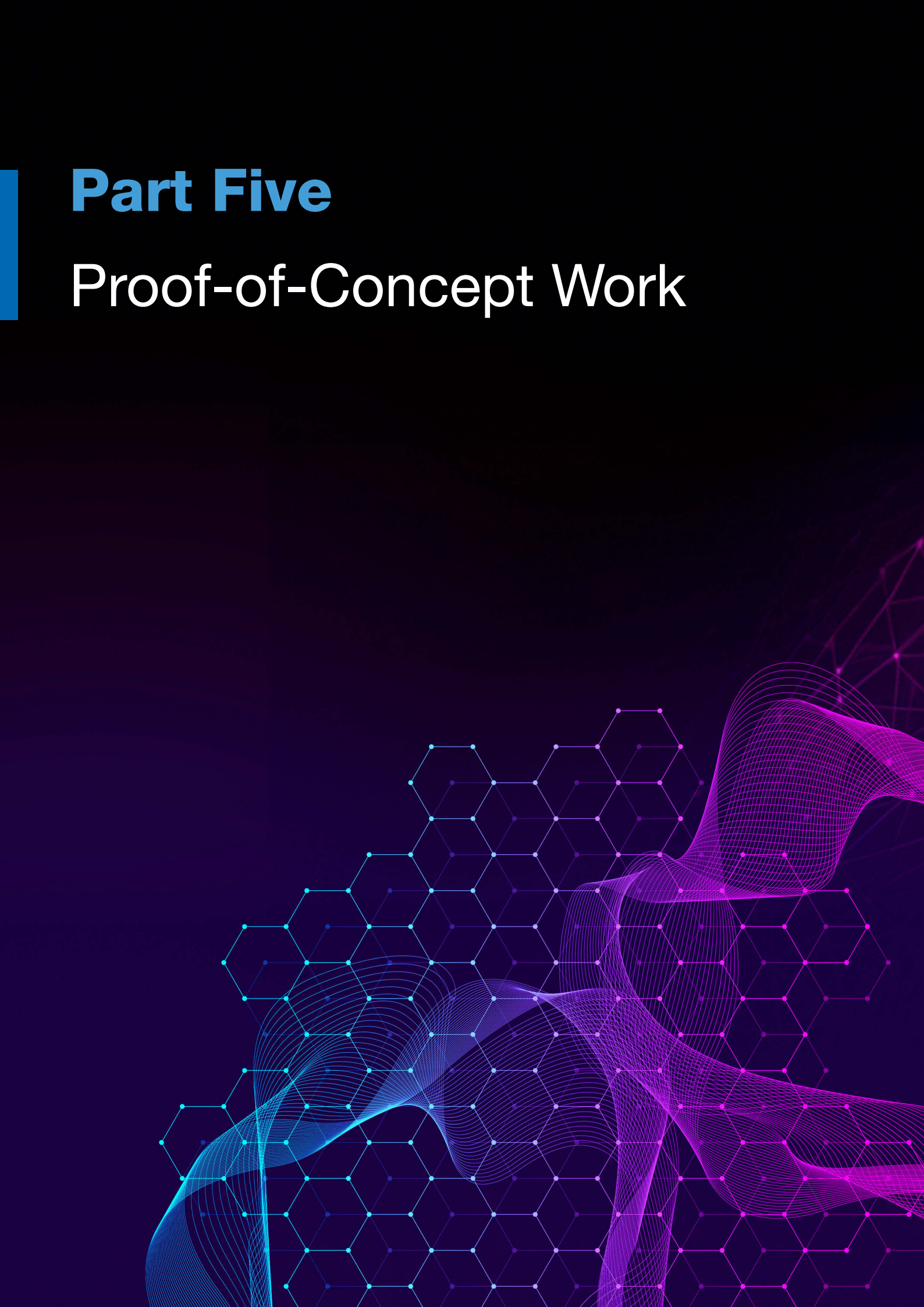
Table 28 Comparison of training times between HE and FTSM across different datasets

Datasets	Existing Approach (Homomorphic Encryption)	Fast-Training Strategy Module (FTSM)
Insurance Dataset on Agency Performance[67]	21 seconds	6 seconds
Prudential Life Insurance Assessment[68]	88 seconds	5 seconds

67 Retrieved from Kaggle: https://www.kaggle.com/datasets/moneystore/agencyperformance
68 Retrieved from Kaggle: https://www.kaggle.com/competitions/prudential-life-insurance-assessment/overview

Part Five

Proof-of-Concept Work



Part Five:
Proof-of-Concept Work

The Proof-of-Concept (PoC) phase during the research focused on three specific use cases that leverage FL to enhance various aspects of insurance operations. These use cases applied three algorithms (Logistic Regression, Boosting, and Neural Networks) to train and test the model. Multiple metrics were used to quantify the performance benefits of the FL approach compared to those resulting from training individual models on separate local datasets. Annex A contains the detailed performance evaluation methodology and results.

In this FL framework, data consumers (e.g. insurers) act as the model training initiator, providing a dataset with prediction labels representing the target outcomes for training, while data providers from various sectors contribute unlabelled, anonymised and encrypted data. Participating parties have the flexibility to collaboratively determine the role of data consumer, thus ensuring that the model's use for prediction aligns with relevant compliance and regulatory requirements. Table 29 below provides an overview of the three use cases along with the roles of the participating parties.

Table 29 An overview of the three Proof-of-Concept use cases

Use case	Purpose	Participating parties
1. Customer Propensity to Purchase	• To enhance the accuracy of the analytical model with engagement insight of customer groups, in order to improve customer targeting and optimise marketing strategies to boost acquisition and retention.	• Insurer A (Data consumer): Provides traditional insurance data, including a label designed to evaluate the propensity of customers to acquire a new policy over a three-month horizon. • Company B (Data provider): Provides anonymised, aggregated and encrypted insights of customer groups of particular demographics with certain behavioural and purchasing attributes.
2. Claim Probability	• To construct a robust predictive model that leverages clinical data to accurately forecast the probability of insurance claims.	• Insurer C (Data consumer): Provides traditional insurance data, including a label indicating whether the customer had ever filed any insurance claim in the past. • Company D (Data provider): Provides historical health data.
3. Renewal Probability	• To integrate comprehensive insurance records with credit rating data to forecast customer renewal probability.	• Insurer E (Data consumer): Provides traditional pet insurance data, including a label used for evaluating the policyholder's overall risk profile, to determine policy renewal. • Company F (Data provider): Provides pet owners' credit rating data.

These use cases were designed to explore the technical feasibility of the proposed FL platform, with the goal of demonstrating how FL can enhance insurance operations. Apart from looking at model performance metrics, the analysis has included examining the business value of each use case, along with implementation considerations, practical challenges, and proposed solutions. Insights have also been derived from qualitative feedback from participating data partners and observations by the research team during the PoC period. This integrated approach has enabled a comprehensive view of the practical applications of FL in the insurance sector to be developed.

5.1 Use Case 1- Customer Propensity to Purchase

5.1.1 Introduction

• Background and motivation

By gaining insights into customers' activities, interests, and purchasing propensity, insurers can tailor their product offerings to better align with customers' needs, resulting in more relevant insurance solutions that enhance customer satisfaction and engagement. However, leveraging alternative data from third parties can raise data security and privacy challenges, especially when personally identifiable information (PII) or other sensitive information is shared. To address these challenges, a life insurance company (Insurer A) collaborated with a company in retail sector (Company B) to research on the development of propensity-to-buy AI models using aggregated data insight from anonymised insurance data and engagement data to conduct model training through FL, enabling Insurer A to better understand the needs of its customers and identify target customers for insurance products.

• Objectives

The primary objectives of the research of this use case were:

1. **Enhanced customer targeting:** To enhance customer targeting and optimise marketing and sales strategies to increase customer acquisition and retention rates.
2. **Improved data analytics:** To leverage integrated data for advanced analytics, enabling insights into customer behaviours and preferences that could inform strategic decision-making across the organisation and enhance prediction accuracy.
3. **Optimised customer support:** To use engagement insights of anonymised customer groups to anticipate customer enquiries, for proactive support and faster resolution times.
4. **Cross-selling opportunities:** To identify potential cross-selling opportunities by better understanding customer behaviours and preferences of customer groups with specific attributes.
5. **Enhanced regulatory compliance:** To strengthen regulatory compliance by implementing robust frameworks for cross-sector data insight research, ensuring adherence to legal standards, and thus fostering trust among stakeholders.
6. **Enhanced knowledge of FL:** To contribute to the existing body of knowledge regarding the adoption of FL in the insurance industry, thereby advancing its application.

5.1.2 Data and Experiments

• Data description

The dataset used originated from two distinct sources. Table 30 below provides a detailed description of the datasets used by Insurer A (the data consumer) and Company B (the data provider). In the course of the research, 1,066 matched rows were identified by the model, with Insurer A's dataset containing 17 features and Company B's having 34 features. Together this makes 51 features per customer, meaning that the dataset is high-dimensional and allows models to capture complex patterns and relationships. While the 1,066 matched rows may seem modest in size compared to large-scale datasets, they nevertheless represent a meaningful sample for predictive modelling in the insurance domain, especially when the data is rich in features.

The model was designed to predict whether a customer would purchase a new policy in the next three months, based on two years of data, with duplicate entries removed to ensure uniqueness. The dataset was divided into training and test sets according to an 8:2 ratio respectively.

Insurer A's dataset included unique identifiers for customers, which are special codes generated from encrypted customer information. These identifiers, referred to as "key', distinguish each data entry while concealing personal details, replacing real data with random-looking strings to ensure privacy and maintain uniqueness. Along with these keys and the target label, the dataset included a variety of features providing a comprehensive view of a variety of attributes for the predictive model. Company B's dataset, consisting of anonymised data, complemented Insurer A's dataset by adding insights on attributes and preferences at aggregated customer group levels.

Table 30 Description of the datasets provided by Insurer A & Company B

Items	Insurer A (Data consumer)	Company B (Data provider)	Prediction Label on Insurer A's dataset
Data types	Traditional insurance data	Alternative data (anonymised and aggregated)	Whether the customer had purchased a new policy within the past 3 months.
Features	• Claim history • Last interaction • Last purchase • Customer since • Residential district • Age group • Astrological sign • Gender • Income range • Industry • Marital status • City of living • Active policy premium • Active policies • Lapsed policies	• Customer engagement tags • Demographic attributes • Purchase behaviour • Redemption behaviour • Payment behaviour • Website browsing behaviour	
Number of rows (matched)	200,000 (1,066)	200,000 (1,066)	
Number of features used	17	34	

• Experimental findings

The federated model results in this research showed a **49.8% improvement** in predictive performance compared to the local models[69], equivalent to approximately 100 more accurate predictions per 1,000 samples. This positive outcome is attributable to the availability of a sufficient amount of diverse and highly relevant data insight. For the detailed experimental results, please refer to section 3.1 in Annex A.

i. Business value unlocked across the value chain

This use case demonstrated a substantial uplift in the accuracy of the Propensity-To-Buy (P2B) model. Three key areas of business value derived from this improvement are laid out below:

- ***Improved targeting and conversion rates***
  With more accurate P2B predictions, an insurer can identify high-intent customers more reliably. This allows marketing teams to focus their efforts on segments most likely to convert, resulting in higher campaign efficiency and increased policy sales.

- ***Optimised marketing spend***
  By reducing outreach to low-probability prospects, an insurer can lower its cost per acquisition. Resources can be reallocated to high-performing channels and personalised campaigns, maximising returns on marketing investment.

- ***Tailored customer engagement***
  The enriched model enables an insurer to develop tailored product recommendations based on customers' engagements and behavioural patterns. Such tailored recommendations can improve customer satisfaction and strengthen long-term customer relationships.

ii. Implementation considerations and limitations

- ***Regulatory compliance***
  In cross-sector data insight research, all parties must navigate various regulatory requirements. Insurance companies must ensure that customer data used in the FL platform, whether operating in the cloud or on local premises, complies with regulatory and governmental standards, particularly the PDPO, and uses only aggregated and anonymised non-PII data, where appropriate.

  In 2023, the IA published the Open API framework to promote data collaboration and connectivity. FL can leverage open APIs to access decentralised data, share model updates, ensure interoperability, monitor performance, and maintain security, thereby enhancing collaboration and maintaining data privacy protection.

- ***Data management and privacy measures***
  Data anonymisation through removal of PII, encryption and aggregation protects individual identities and maintains privacy, while data minimisation principles ensure that only necessary data insight is collected and used in the research. Together, these measures enhance compliance with regulatory standards and strengthen overall data security and protection.

- ***Access controls***
  Strict access controls safeguard data and ensure that only authorised personnel can access the datasets used in the research and the output data, thus maintaining data integrity and security.

- ***In-house knowledge and skills required***
  Successful implementation requires full engagement from the business development team. A full-stack data expert is needed to monitor the ML model's performance and quantify the generated business value. Additionally, a thorough understanding of legal and regulatory requirements related to data privacy and security is essential.

- ***Dataset limitations in the POC research project***
  As a POC research project, there were inherent limitations regarding the dataset, including its size and diversity, which may have impacted the robustness of the findings.

69 The Ratio of Improvement of Federated Learning (RIFL) is a metric used to quantify the performance benefits of the FL approach compared to training individual models on separate local datasets. For more details of the performance evaluation methodology, please refer to Annex A section 1.

iii. Challenges and solutions

◇ ***Challenges related to data quality and availability***

a. Matched Sample Data Deficiencies and Data Completeness

A key challenge in this use case of this research was the lack of matched sample data, and data completeness issues. To tackle this, the training and testing datasets were temporarily combined to increase the sample size for model development. While this helped mitigate the impact of incomplete data, it also risked limiting the ability to evaluate model generalisations. To counter this, additional validation strategies, such as cross-validation, were employed.

Additionally, low quality fields were removed from the datasets, improving data integrity and ensuring the relevance and reliability of the remaining dataset.

b. Insufficient Data Processing Before Training

Another challenge encountered was insufficient data processing prior to training the model. Incomplete or poorly processed data can lead to suboptimal model performance, as the quality of input data directly affects the accuracy and reliability of predictions. Thorough data preprocessing, such as data cleaning and outlier detection, must be carried out before data is uploaded to the platform.

c. Risks of Bias and Inaccuracy from Missing Values

Missing values in the large dataset significantly increase the risks of model bias and inaccuracies, wasting computational resources. To address this, the solution simplified the ML model by removing data fields with substantial missing values.

◇ ***Data privacy and security risks***

Collecting and using Personally Identifiable Information (PII) in data analytics brings privacy challenges. To address these challenges, a unique identifier system is implemented on the datasets of Insurer A (data consumer) to replace real data with random-looking strings and concealing personal details. For the dataset of Company B (data provider), all PII were removed from the dataset before they were used for the research. These measures enhanced security during the matching process and prevented unauthorised personnel from inferring meaningful information from the raw data.

Additional measures adopted included:

- **Local model training and storage:** Throughout the research process, all datasets remain securely within the premises of their respective owners—data providers and data consumers. Only encrypted model updates are shared for their respective on-site deployment, reducing privacy risks and enhancing protection against data breaches.

- **Homomorphic encryption:** AES-256 (Advanced Encryption Standard with a 256-bit key) secured data by converting it into an irreversible format, allowing encrypted data to be processed during model training and throughout the research process without decryption.

- **Data erasure:** All output data generated from the model was erased from the relevant platform after the completion of the research for better privacy protection.

- **Rigorous governance and review processes:** Privacy Impact Assessments (PIA) and Information Security Risk Assessments (ISR) were conducted to ensure compliance and data security.

◇ ***Platform constraints***

The FL platform supports only limited model types and tuning options, making it less flexible. More training rounds and manual tuning can help improve performance, but this takes a lot of time, especially when training is spread across many locations. In production, MLOps tools can solve this by automating testing, tracking, and training, making updates easier to manage and scale.

◇ ***Network and compliance issues***

Data consumer and data provider often operate on different network infrastructures. Moreover, security requirements and compliance procedures required by both data consumers and data providers can result in a lengthy approval process.

To address these challenges, the stakeholders established effective communication channels for proactive discussions on compliance and security measures. They engaged legal and compliance teams early in the planning stage to obtain professional advice and review engagement documentation, while also developing a standardised protocol to streamline the information exchange between diverse network infrastructures. To further address data privacy and security concerns, a tri-party non-disclosure agreement was signed along with a license evaluation agreement specifically for this POC project.

◇ ***Model selection difficulty.***
Identifying the best model for any given task is challenging, as model performance is heavily influenced by the characteristics of the data used for training and evaluation. Factors such as dataset size, diversity, and feature distribution all determine a model's effectiveness for particular applications.

To overcome this challenge, stakeholders should conduct data profiling before model selection, ensure their model choice is in alignment with their particular business objectives, and use domain-specific evaluation metrics to validate the model performance.

iv. Objective evaluation

The primary objectives of this use case under the research project were successfully achieved. The research results demonstrate that improved data analytics significantly boost predictive power, providing deeper insights into customers' behaviours and preferences and leading to better customer targeting, cross-selling opportunities, and customer support.

The FL platform's privacy-preserving architecture, supported by robust data protection measures implemented by data providers and consumers, adhere to established privacy standards for data protection and secured information exchange. Nevertheless, additional measures such as regular audits and robust incident response plans are needed to fully ensure compliance and foster greater user trust.

This use case has deepened the understanding of FL among the insurance industry, providing insurers with valuable firsthand experience of its applications. Overall, it demonstrates FL's ability to leverage engagement insights of anonymised customer groups to enhance Propensity-To-Buy models, offering a privacy-conscious approach to more effective customer targeting for the insurance sector.

5.2 Use Case 2- Claim Probability

5.2.1 Introduction

• Background and motivation

Insurance companies closely monitor the number and cost of claims they receive, as this information is crucial for maintaining financial stability, managing risk, and ensuring long-term profitability.

Given this, Insurer C collaborated with Company D (a one-stop healthcare centre) on an FL system that can develop robust claim models while respecting privacy. This use case aimed to evaluate how the addition of clinical data could affect the accuracy of insurance claim predictions.

• Objectives

1. **Predictive model development:** To develop a predictive model for forecasting insurance claims by integrating insurance and clinical data.

2. **Impact analysis of clinical measurements:** To evaluate how policyholders' clinical data affects their likelihood of making insurance claims. With health insurance becoming increasingly important in Hong Kong, this use case sought to provide tools and insights for better risk management within the industry.

3. **Collaboration opportunities:** To foster a partnership between an insurer and a healthcare provider by leveraging shared insights and data-driven strategies to enhance patient care and health outcomes.

4. **Regulatory compliance:** To ensure compliance with privacy regulations by using FL to enable secure, compliant data sharing and processing.

5. **Enhanced decision-making:** To utilise predictive analytics to facilitate informed decision-making in underwriting and claims management, while also developing targeted risk mitigation strategies to reduce the likelihood of high-cost claims and enhance overall portfolio performance.

5.2.2 Data and Experiments

• Data description

Table 31 provides an overview of the dataset resulting from the collaboration between Insurer C (the data consumer) and Company D (the data provider). The dataset comprised 1,000 rows from Insurer C and 403 from Company D. After confidential identity matching, 312 rows were matched, and the resulting dataset was split into an 8:2 ratio for training and validation respectively.

The dataset combined traditional insurance data from Insurer C with health data from Company D, excluding any missing entries. The number of features listed in Table 31 reflects those used throughout the model training cycle, with the prediction label focusing on whether customers had filed insurance claims.

Table 31 Description of the datasets provided by Insurer C and Company D

Items	Insurer C (Data consumer)	Company D (Data provider)	Prediction Label
Data types	Traditional Insurance data	Historical health data	Whether the customer had ever filed any insurance claim in the past
Features	• Age • Gender • Policy month	• Clinical measurements of patients (height, weight, fat, metabolic age, etc) • Biomarkers of patients (cholesterol level, fasting blood glucose, etc)	
Number of rows (matched)	1,000 (312)	403 (312)	
Number of features used	3	36	

• Experimental findings

The results demonstrated that incorporating alternative data, particularly historical health data, significantly enhances the performance of predictive models, achieving a performance improvement **two times better** than when utilising only traditional data. Additionally, a notable and unexpected finding from this use case is that men, despite generally engaging in riskier behaviours, exhibited a lower likelihood of filing claims. This insight could prompt insurers to reassess their pricing strategies, product offerings and customer engagement approaches to ensure greater fairness, accuracy, and service effectiveness. For the detailed experimental results, please refer to section 3.2 in Annex A.

i. Business value unlocked across the value chain

This use case demonstrated a notable improvement in the accuracy of insurance claim prediction models. The areas of potential business value derived from this enhancement are laid out below, highlighting tangible benefits across underwriting, pricing, and product development.

◇ ***Improved risk assessment***

More accurate predictions in claims probability help insurers assess risk more precisely, leading to better underwriting decisions and reduced loss ratios.

- ***Optimised pricing strategies***
  Enhanced model performance supports more tailored pricing, improving competitiveness and profitability.

- ***Product innovation***
  Insights from health data can inform the design of new insurance products that better meet customer needs, especially in the health and wellness segments.

- ***Operational efficiency***
  Better predictions reduce manual reviews and claim processing time, lowering operational costs.

ii. Implementation considerations and limitations

Use Case 2 shares several implementation considerations and limitations with Use Case 1, including the need to comply with data protection regulations, safeguard privacy through encryption and anonymisation, effectively manage diverse network infrastructures, and maintain strong in-house expertise to ensure secure and successful deployment.

However, Use Case 2 also revealed some additional considerations and limitations, including:

- ***Data quantity and quality***
  The accuracy of the model's predictions depends on the integrity and comprehensiveness of the input data. In this use case, the dataset was limited to only 312 samples, presenting a constraint for training robust machine learning models. Additionally, the dataset with fewer features was used to reduce complexity and cost, with Insurer C providing 3 features and Company D 36—an unequal distribution that posed challenges in model training and performance evaluation. Given the small sample size and feature imbalance, it is important to consider traditional statistical techniques as a baseline, as they are often more suitable and interpretable under such constraints. Ultimately, ensuring a sufficient quantity of high-quality data and balanced features is essential for achieving meaningful results in FL model training.

- ***Risk mitigation***
  While the model is designed to mitigate risks by predicting claim probabilities, it cannot eliminate these risks entirely. Insurance claims are influenced by many factors, some of which may be unforeseen or difficult to quantify. Therefore, the model should be viewed as a risk management tool rather than a definitive predictor of future claims.

iii. Challenges and solutions

- ***Heterogeneous data sources***
  While larger and more diverse datasets may improve model performance, they also increase complexity and require careful data standardisation and preprocessing, which can be time-consuming and labour-intensive. Close partnership and good communication between Insurer C and Company D helped solve these challenges.

- ***Operational burden during model iteration***
  The process of model iteration in this use case posed notable operational challenges, particularly due to the need for frequent dataset revisions and renewed consent from data providers. Proactive planning and efficient workflows can effectively address these issues. Close collaboration between data partners and the implementation of robust data governance processes can minimise the need for frequent dataset revisions and consent renewals. Additionally, automated tools and techniques can streamline the model iteration process, reducing the overall operational burden. For instance, robotic process automation such as UiPath can automate repetitive tasks, allowing users to integrate them seamlessly into the platform.

- ***Adapt to diverse network infrastructures***
  In Use Case 2, the fact that the data partners had different network infrastructures introduced unexpected technical challenges. These included issues relating to network proxies and inconsistent bandwidth, which affected the efficiency of model updates and parameter exchanges. To address these issues, the FL solution must be flexible enough to accommodate the distinct network conditions and varied network infrastructures of each data partner.

  Utilising lightweight, low-overhead protocols, along with techniques like data compression, batching, and adaptive transmission rates, can reduce the network strain arising from frequent model updates and parameter exchanges. Additionally, intelligent monitoring and dynamic protocol selection based on real-time network conditions can ensure the FL process remains efficient and resilient to fluctuations in network performance across diverse infrastructures.

◇ ***Data privacy and security risks***

To mitigate potential data privacy and security risks associated with FL, the data partners implemented several control measures:

- **Data encryption:** Sensitive data was encrypted both at rest and in transit. Participants used AES-256 for data at rest and TLS protocols for data in transit, ensuring any intercepted data would be unreadable by unauthorised parties.

- **Access controls:** Strict role-based access controls were established to limit data access to authorised personnel only. Multi-factor authentication (MFA) added an extra layer of security, reducing the risk of unauthorised access.

- **Data minimisation and anonymisation:** The data participants adhered to the principle of data minimisation by sharing only essential information for the PoC. Personal identifiers were avoided or replaced with anonymised IDs to protect individual identities.

- **Secure infrastructure:** The PoC environment was hosted on secure servers with hardened operating systems and up-to-date security patches. Firewalls and intrusion detection systems guarded against external threats.

- **Employee training:** All team members received regular training on data security best practices, confidentiality obligations, and protocols for handling personal and health information, reducing the risk of human error and insider threats.

- **Incident response plan:** An incident response plan was in place to address potential data breaches promptly. It included procedures for containment, eradication, recovery, and communication with affected parties and regulators.

- **Data retention and destruction:** Data was retained only as long as necessary for the PoC. Upon completion, all data was securely destroyed using industry-compliant methods to prevent data reconstruction.

iv. Objective evaluation

The primary objectives of this use case were successfully achieved. A predictive model integrating insurance and clinical data was developed that proved effective in forecasting insurance claims. It showed significant performance improvement compared to traditional models.

Furthermore, the analysis of clinical data uncovered valuable insights that could help in developing innovative products tailored to customer needs. This use case indicates that strong partnerships between the insurance and health sectors have the potential to drive data-driven approaches that improve patient care and health outcomes. Finally, regulatory compliance was effectively maintained using FL, facilitating secure cross-sector data collaboration while adhering to privacy regulations. Overall, these outcomes highlight the transformative potential of integrating health data in the insurance industry.

5.3 Use Case 3- Renewal Probability

5.3.1 Introduction

• Background and motivation

Insurers need to be able to accurately predict renewal probability in order to identify at-risk policies and effectively plan customer retention strategies. In this use case, Insurer E leveraged credit data from Company F to predict policy renewals for their pet insurance offerings. Incorporating credit information is uncommon in the insurance industry in Hong Kong, making this collaboration a pioneering effort that could set a precedent for other insurers.

• Objectives

1. **Enhanced predictive model:** To develop an accurate predictive model utilising advanced FL techniques to reliably predict customers' renewal probabilities for insurance policies.

2. **Safeguard customer privacy and compliance:** To ensure robust protection of customer data and compliance with all relevant data privacy regulations throughout the modelling process.

3. **Enhance customer insights:** To leverage FL and alternative data sources (e.g. credit data) to uncover insights into customer behaviour and preferences, identifying opportunities for tailored marketing strategies and product offerings that boost customer engagement.

4. **Foster strategic partnerships:** To establish collaborations with data providers, such as Company F, to enrich available data sources. This use case partnership aimed to gain insights from the integration of credit data and pet insurance renewal data.

5. **Drive innovation in the market:** To leverage advanced analytics and credit data to uncover insights that can drive innovation and foster a culture of continuous improvement.

5.3.2 Data and Experiments

Table 32 provides an overview of the dataset used in this use case. The credit data from Company F was synthetic testing data, generated to emulate real data for this use case. The dataset employed the encrypted HKID of the customer as the unique identity, with the label of evaluating the policyholder's overall risk profile to determine policy renewal decisions.

The total number of rows in the dataset was 2,000 from Insurer E and 19,201 from Company F. After confidential identity matching, 1,957 rows were matched, and these matched rows used to participate in the training and back test phases. The whole matched dataset was split into an 8:2 ratio for model training and validation respectively.

Table 32 Description of the datasets provided by Insurer E and Company F

Items	Insurer E (Data consumer)	Company F (Data provider)	Prediction Label
Data types	Traditional pet insurance data	Pet owners' credit rating data	Whether the customer had renewed the policy.
Features	**Historical claim data:** • Age of pet • Purchased product count • Purchased policy count • Average premium of issued policies • Claim amount • Claim ratio • Purchase flow	• Payment history • Account status credit exposure • Load amount credit limit • Credit utilisation • Product holding (e.g. Type, number, credit length, new credit) • Enquiry footprint • Credit score • Public records	
Number of rows (matched)	2,000 (1,957)	19,201 (1,957)	
Number of features used	10	10	

Company F unexpectedly withdrew from the POC project before it was completed due to the cessation of its operations following a depletion of funds. This early exit limited the scope of collaboration and experimentation.

• Experimental findings

Due to the exit of Company F, Insurer E was unable to thoroughly test and fine-tune data features to achieve optimal results. Additionally, some basic data preparation issues could not be addressed, such as the substantial amount of data missing from Company F's dataset, which prevented Insurer E from achieving a comprehensive understanding of the data landscape. This use case once again highlights the importance of data quality and communication among data partners in the context of FL.

i. Business value unlocked across the value chain

The use case sought to enhance predictive modelling through advanced FL techniques in order to unlock significant business value potential for Insurer E. The unexpected exit of Company F hindered thorough testing and optimisation of the predictive accuracy for customer renewal probabilities. However, the use case did underscore the importance of data-driven decision-making, as data privacy was upheld. This experience emphasises the need to prioritise data quality and effective communication among cross-sector data partners.

ii. Implementation considerations and limitations

◇ ***Regulatory compliance***

Cross-sector data collaboration involving credit data in Hong Kong is governed by strict regulations, such as the PDPO. It mandates that organisations collect only necessary personal data, obtain express and voluntary consent if the data were to be used for a new purpose which is not or is unrelated to the original purpose upon collection, and implement robust security measures such as encryption and access controls. Violations may result in significant penalties.

The HKMA imposes additional requirements through its Supervisory Policy Manual, specifically the module on "The Sharing and Use of Consumer Credit Data through Credit Reference Agencies". This involves proper governance, data accuracy, and consumer protection when sharing or using credit data by authorized institutions.

Building consumer trust requires transparency about data usage and confidentiality agreements. It is recommended that secure communication protocols for data transmission between devices and the central server are established, and privacy-preserving techniques such as differential privacy and secure multiparty computation are employed.

◇ ***Data quality and preprocessing***

This use case was characterised by missing data, a common issue in financial datasets which can have significant implications. For instance, missing data in an income field may suggest that an applicant is self-employed or has inconsistent income sources, impacting creditworthiness assessments. Since data quality directly affects model performance, the data must be thoroughly preprocessed before training.

Key data preprocessing tasks include:

- Handling missing data or outliers
- Performing feature engineering and selection
- Normalising and scaling data variables
- Addressing class imbalances in the dataset

◇ ***Regular monitoring and evaluation of model performance***

In this use case, the model locally trained by the data consumer demonstrated satisfactory performance, suggesting that renewal probabilities could be effectively captured by the local data and features. Consequently, integrating additional data sources like credit data may not necessarily improve the model performance, and could even degrade it.

Therefore, it is important for the FL platform to allow data consumers to train their models locally with ML algorithms.

iii. Challenges and solutions

◇ ***Data related issues***

In this use case, the random selection of datasets by the data partners led to numerous missing fields, resulting in inaccurate model performance. Additionally, a lack of familiarity with each other's data fields made it challenging to interpret the results. Data partners must clean their data to ensure relevance and accuracy before uploading it to the platform.

Furthermore, prior to model training, data partners should engage in thorough discussions and planning regarding data fields and structure to ensure consistent definitions. Maintaining close communication with the system developer is also essential for understanding system limitations and the uploading process.

◇ ***Data security and privacy risks***

To mitigate the security and privacy risks associated with FL, Insurer E implemented several control measures:

– **Minimal data sharing:** Only HKID numbers, a form of Personally Identifiable Information (PII), were included in the data to minimise the risk of exposing sensitive information.

– **Encryption of sensitive identifiers:** HKID numbers were encrypted using the SHA-256 hashing algorithm to enable secure data matching and analysis within the FL model without disclosing original HKID values. While SHA-256 is a widely adopted cryptographic method, it is generally recommended to apply additional safeguards to enhance protection against potential reverse techniques.

– **Data localisation:** All raw data remained within the organisation's infrastructure, with no transfer to the data partner. This practice eliminated the risk of data exposure during the training process.

◇ ***Data provider exited during the POC***

Effective model training typically necessitates multiple rounds of testing and fine-tuning for optimal results. Unfortunately, the data provider in this study exited the market after the initial training round due to unforeseen circumstances. This hindered the ability to conduct further training to refine and enhance the accuracy of the model. To address this, experiments were simulated to verify the findings.

◇ ***Constraints of leveraging additional data***

The use case also revealed that model performance is not necessarily enhanced by adding additional data in cases of poor data quality, insufficient data quantity, or irrelevant features. It underscored the importance of carefully evaluating data quantity, quality and relevance prior to model training.

◇ ***Resource-intensive training process***

The time and computation power required for the training process to be completed, especially for techniques like Boosting, can be considerable, so all participating devices must have sufficient computational power. Allocating computational and network resources effectively across devices helps in maintaining a balanced workload and optimal performance.

Table 33 summarises the challenges encountered in this use case, including reasons, risks and mitigation strategies.

Table 33 Summary of the challenges, potential risks and mitigation strategy

Challenges	Potential Reason	Potential Risk	Mitigation Strategy
Missing value in datasets	• Uncleaned dataset • Incomplete data collection • Data entry errors	• Reduced accuracy of the model due to lack of information • Model bias due to reliance on incomplete data	• Implement imputation techniques (mean, median, mode) or one-hot encoding to fill in missing values • Conduct data preprocessing
Unpredictable performance improvements from additional data	• Less relevant dataset is provided • Data consumer already has high-quality data	• Slows down the convergence of the federated model • Time wasted in training useless models	• Use auto selection to remove unused features from each model • Small batch pre-training to filter out the most relevant features
Resource-intensive training process	• Boosting calculation is not optimised • Secrete sharing takes up lots of memories	• Long training time may crash the system • Hyperparameter optimisation becomes impossible due to limited resources	• Select fewer columns for training or use a larger validation ratio for Boosting • Small batch pre-training to filter only the most useful features
Difficulty in interpreting the model results	• Lack of understanding of data fields by data partners	• Inability to derive meaningful insights from model results	• Engage in comprehensive discussions about data definitions, structures, and context before training • Establish a common framework and glossary of terms to enhance clarity among partners

iv. Objective evaluation

While the objective of enhancing the predictive model was not fully realised due to the unexpected exit of the data provider, simulated experiments indicated that model performance could significantly improve if the missing data issues are addressed and multiple rounds of fine-tuning conducted.

The objective of safeguarding customer privacy and ensuring compliance with data regulations was successfully met. Furthermore, given that the integration of credit data into the insurance sector is relatively uncommon in Hong Kong, this collaboration could be a valuable reference point for future initiatives. Other insurers in the region could emulate this data collaboration model by developing similar partnerships.

5.4. Conclusion

5.4.1 Key Insights

The PoC cases demonstrated the potential of FL for the insurance sector in the following three areas:

1. Smarter predictive models

a) Proven effectiveness of FL

FL showed its ability to develop improved predictive models in most use cases through the incorporation of alternative data, as shown by the fact that the federated models performed better than the local models. Being able to more accurately predict claim probabilities and customer propensity to purchase enables insurance companies to undertake more precise pricing of their insurance policies, optimise their resource allocation, and make their marketing strategies more effective.

In addition to the use cases explored in this research, FL shows strong potential for broader applications across the insurance value chain, as indicated by prior studies and industry initiatives. By enabling industry-wide data collaboration, FL allows insurers to jointly train machine learning models while preserving the confidentiality of proprietary and customer information. Key areas where this approach is particularly promising include:

- **Risk assessment:** Collaborative modelling of underwriting risks and expected losses enhances predictive accuracy without compromising data privacy.

- **Pricing optimization:** Shared market insights support refined pricing strategies, enabling competitive positioning while safeguarding sensitive pricing structures.

- **Fraud detection:** Cross-insurer model training helps identify organized or syndicated fraud, fostering shared intelligence without exchanging raw claims data.

- **Customer behaviour analytics:** Analysing integrated data across functional domains (e.g. claims, transactions, customer interactions, and operational records) to identify trends and provide useful insights that allow for personalised services while maintaining data sovereignty and compliance.

Table 34 shows that its advantages are being realised in diverse areas in the financial sector, such as marketing, pricing, risk management, and fraud detection. These examples highlight FL's growing strategic importance for the insurance industry, where secure, data-driven collaboration is a key driver of innovation and competitive advantage.

Table 34 Proved benefits of FL to the financial sector

Application	Key Impacts	Performance Gain	Collaborators	Source
Marketing and Sales	• Improved cross-selling conversion in bancassurance	• Over 50%	• Insurers • Banks	• China Academy of Information and Communications Technology (2022)[70]
Pricing	• personalised pricing coverage expansion	• 10% ➔ 92%, 1.5 times increase in profits	• Insurers • Internet companies	• WeBank[71]
Risk Management	• Improvement in SME loan risk control model	• 12%	• Banks • Collaborative companies	• WeBank
Fraud Detection	• Improvement in cross-institutional fraud detection	• 30%	• Banks • Data providers	• China Academy of Information and Communications Technology (2022)

b) Importance of data quality and diversity

As with any data-driven approach, the success of FL depends on the quantity, quality, and diversity of the input data. The PoC emphasised the need for:

- High-quality, relevant, and diverse datasets
- Consistent data formats and structures across partners
- Thorough data preprocessing and validation

c) Strategic resource allocation

Allocating adequate resources, such as sufficient computation power, is also crucial for supporting comprehensive training and swiftly addressing technical and communication issues that arise during the process.

d) Tailored algorithm selection

Different algorithms have varying degrees of predictive capability, and data scientists or machine learning experts need to be available to select the most suitable algorithms.

2. Collaborative cross-sector partnerships

Insurers often struggle to build robust predictive models due to limited customer interactions, resulting in a lack of labelled data. FL addresses this by enabling insurers to collaborate with other organisations to jointly develop ML models without sharing sensitive data. This approach helps reduce legal and operational barriers to data collaboration while supporting the development of new insights, such as patterns in claims, fraud, and customer behaviour across institutions. While FL has the potential to alleviate compliance challenges in data collaboration, strong partnerships and a high level of trust are essential.

3. Enhanced data privacy

FL enables insurance companies to train ML models on decentralised data without disclosing sensitive customer information, supporting strong data privacy practices and compliance with stringent data protection regulations. As the regulatory landscape continues to evolve, FL presents a valuable opportunity for insurers to align with emerging

70 中國信息通信研究院泰爾終端實驗室, 聯邦學習場景應用研究報告, 2022.
71 陳天健, 基於聯邦學習新技術連接數據孤島, accessed 7 August 2025, https://static001.geekbang.org/con/40/pdf/2790523233/file/陳天健-基於聯邦學習新技術連接數據孤島.pdf.

standards and best practices. By strengthening their internal compliance capabilities, fostering proactive engagement with regulators, and adopting advanced privacy-preserving techniques, insurers can further support the responsible and effective deployment of FL across the sector.

5.4.2 Recommendations for Effective Implementation

Based on the experience of the PoC and feedback from participants, we recommend the following to facilitate effective FL implementation:

• Start with a pilot and a specific use case

Implementing FL is a complex undertaking, particularly in sectors like insurance where its applications are still emerging. Starting with a well-defined pilot allows an organisation to minimise risks and build a solid foundation for broader FL implementation. Here are some key considerations:

- **Pilot phase:** Begin with a focused pilot project to test the FL model in a controlled setting.
- **Feasibility testing:** Assess the model's operational fit, including technical requirements and data availability.
- **Challenge identification:** Use the pilot to uncover potential issues, such as data privacy concerns and system integration needs.
- **Communication channels:** Establish clear communications with partners to facilitate collaboration and alignment on goals.
- **Refinement:** Collect feedback to make necessary adjustments, enhancing methodologies and data practices before scaling up.

• Ensure data quality, compatibility and availability

High-quality data that is accurate, complete, and representative of the target problem is a key priority. Organisations should implement standardised data quality validation protocols, such as ISO/IEC 25012[72] and ISO/IEC 25024[73], to ensure data quality and compatibility across various data sources. Additionally, to comply with PDPO while ensuring data availability for FL, for example, organisations shall inform customers about data usage through a Personal Information Collection Statement (PICS). Insurers can also follow PCPD's recommendations to consider anonymising personal data in accordance with the recommended steps as stated in the "Guide to Getting Started with Anonymisation"[74].

In Hong Kong, the data governance framework is guided by the Principles of Data Governance[75] introduced by the Digital Policy Office (DPO) in December 2024. The government and related organisations also provide technical standards, including the IT Security Standards and Best Practices[76] and the Ethical Artificial Intelligence Framework[77], for both public and private institutions. Moreover, when utilising generative AI technology, developers, service providers, and users should refer to the Hong Kong Generative Artificial Intelligence (AI) Technical & Application Guideline released by the DPO in April 2025. This guideline covers essential areas including the scope and limitations of generative AI applications, governance principles, and potential risks such as data leakage, model bias, and system errors.

Organisations utilising geospatial data must adhere to established standards like ISO 19157[78] for data quality assessment and ISO 19115[79] for interoperability. These protocols enable organisations to minimise errors during data entry and processing, ensuring the reliability and integrity of their geospatial information while complying with regulations such as GDPR and maintaining secure data storage solutions.

---

72 ISO/IEC 25012 is an international standard that defines a general data quality model for data retained in a structured format within a computer system. It can be used to establish data quality requirements, define data quality measures, or plan and perform data quality evaluations.

73 ISO/IEC 25024 is an international standard that defines data quality measures for quantitatively measuring the data quality in terms of characteristics defined in ISO/IEC 25012. It can be applied to any kind of data retained in a structured format within a computer system used for any kinds of applications.

74 Asia Pacific Privacy Authorities ("APPA"), Guide To Getting Started with Anonymisation, June 2025.

75 In December 2024, the Digital Policy Office (DPO) launched a thematic web page on data governance, providing a one-stop resource for the government's data governance policies. This page includes the Principles of Data Governance, relevant strategies, guidelines, and technical standards. https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/policy/

76 IT Security Standards and Best Practices refer to a set of internationally recognized guidelines and frameworks designed to help organisations manage information security effectively.

77 The Ethical AI Framework, developed for internal use within the Hong Kong Government, guides the ethical application of AI and big data analytics in IT projects. It assists bureaux and departments in incorporating ethical principles and practices during planning, design, and implementation. This framework, including its guiding principles and assessment templates, has been revised for broader applicability, allowing other organisations to use it as a reference when adopting AI and big data analytics in their projects. https://www.digitalpolicy.gov.hk/en/our_work/data_governance/policies_standards/ethical_ai_framework/

78 ISO 19157 is an international standard that provides guidelines for assessing the quality of geographic data, focusing on metrics such as accuracy, completeness, and consistency. It aims to ensure reliable and trustworthy geospatial information for users.

79 ISO 19115 is an international standard that specifies the metadata schema for describing geographic information and services, enhancing data discoverability and interoperability. It provides guidelines for documenting data quality, ensuring effective management and sharing of geospatial datasets.

• Foster strong partnerships

Effective implementation of FL requires close collaboration between solution developers and project partners. Key elements for strong partnerships include:

- **Establish clear objectives:** Define specific goals to guide the partnership, and ensure all partners understand their roles.
- **Establish a partnership agreement:** Establish a Non-Disclosure Agreement (NDA) or formal partnership agreement to outline the terms of collaboration, such as the roles and responsibilities of each partner, data ownership, intellectual property rights, confidentiality considerations, compliance and ethical considerations, and termination conditions.
- **Build trust through transparency:** Maintain open communication about project developments and data usage to foster confidence among partners.
- **Utilise secure communication channels:** Use encrypted messaging and secure data sharing platforms to protect sensitive information.
- **Conduct regular feedback sessions and meetings:** Schedule frequent discussions and meetings to review progress, share insights, and address challenges early, strengthening relationships and reinforcing commitment to common goals.

• Develop robust privacy protocols

The FL solution must have robust security measures, such as strong access controls, encryption, and secure communication channels. Constant vigilance and regular security audits are crucial to identify and address any vulnerabilities in the solution. Certifications like ISO/IEC 27001[80] demonstrate a commitment to security standards and enhance stakeholder trust.

Moreover, legal and compliance teams should ensure that the FL initiative adheres to relevant data protection regulations, such as the PDPO, particularly section 4 and the DPPs, which outlines the principles of data protection, including the requirement for personal data to be collected fairly and lawfully, and the use of personal data be limited to or related to the original collection purposes only. When using cloud computing to process personally identifiable information (PII), users should also comply with ISO/IEC 27018[81] by implementing strong data protection measures, obtaining explicit user consent, and maintaining transparency about data usage. Regular audits and risk assessments are also essential for ongoing compliance and building trust.

• Adopt a comprehensive platform

Successful FL implementation requires integrated solutions that manage the entire lifecycle effectively, from data preprocessing to model training and result analysis. Key components include:

- **Data preprocessing:** Utilise platforms that automate data cleaning and anonymisation.
- **Model training:** Support distributed training across nodes, allowing local data processing.
- **Result analysis:** Implement advanced analytical features for performance insights.
- **Collaboration tools:** Enable seamless communication among stakeholders.
- **Scalability and flexibility:** Choose solutions that can scale and adapt to various use cases.

Prior to deployment, organisations should conduct external security assessments, such as the Security Risk Assessments and Security Audits (SRAA) and Privacy Impact Assessments (PIA), to identify potential security vulnerabilities and privacy risks associated with sharing and processing data across platforms. Furthermore, organisations should align their security practices with internationally recognised frameworks such as ISO/IEC 27001, which provides standards for information security management systems, and ISACA's COBIT framework, which offers guidelines for governance and management of enterprise IT. These frameworks ensure robust security measures and effective risk management when implementing FL systems.

80 ISO/IEC 27001, known more commonly as ISO 27001, is the leading globally recognized information security standard, developed jointly by the International Organization for Standardisation (ISO) and the International Electrotechnical Commission (IEC). This certification focuses on information security management systems (ISMS) and is crucial for ensuring the confidentiality, integrity, and availability of data.

81 ISO/IEC 27018 is an international standard that provides guidelines for protecting personally identifiable information (PII) in cloud computing. It specifies controls and security measures that cloud service providers must put in place to protect their customers' personal data.

• Ensure infrastructure readiness

Training FL models can be computationally intensive, and organisations will need robust infrastructure capable of supporting larger volumes and more diverse datasets. This necessitates significant investment in IT infrastructure to ensure seamless operations as data volumes grow.

To ensure infrastructure readiness for a FL platform, organisations should:

- Assess their current IT capabilities for handling increased data.
- Upgrade to high-performance servers and storage.
- Consider scalable cloud and distributed computing solutions.
- Ensure a robust network for fast data transfer.
- Establish efficient data management practices.
- Set up monitoring tools and schedule regular maintenance.
- Implement strong security measures to protect data.
- Provide training and ongoing support for staff.

• Ensure interoperability with existing systems

Integrating FL with legacy systems requires careful planning in the following areas:

- ◇ **Assessment of legacy systems:** Evaluate existing systems to identify compatibility issues.
- ◇ **API development:** Create APIs that enable FL systems to communicate with legacy applications.
- ◇ **Data standardisation:** Standardise data formats across systems.
- ◇ **Gradual integration:** Adopt a phased approach, starting with pilot projects.
- ◇ **Training and support:** Provide training on using the new FL framework alongside existing systems.
- ◇ **Monitoring and maintenance:** Establish processes for ongoing monitoring.

• Translate FL improvements into business value

To effectively translate FL improvements into business value, several key elements should be focused on.

- ◇ **Enhanced data privacy:** FL mitigates compliance risks and reduces potential costs from data breaches, fostering greater stakeholder trust.
- ◇ **Operational cost analysis:** A thorough examination reveals significant resource savings, showcasing FL's efficiency in leveraging decentralised data.
- ◇ **Case studies:** Relevant examples illustrate FL's real-time adaptability for quicker market responses.
- ◇ **Modelling scalability:** FL shows potential for market expansion with minimal investment.
- ◇ **Performance metrics:** Presenting metrics that highlight accuracy improvements, along with discussions on reduced maintenance costs and new collaboration opportunities, underscores how FL enhances model performance while driving business growth and customer satisfaction.

5.4.3 Future Enhancements

The PoC findings reveal that the FL platform utilised represents a significant advancement in collaborative ML, particularly for insurance companies seeking to leverage alternative data sources without compromising data security and privacy.

To improve the effectiveness and capabilities of the platform, the following developments are proposed for future study and application:

- **Increase upload data size limit:** Extend the supported upload data size limit.
- **Expand data preprocessing techniques:** Provide more comprehensive data preprocessing options to improve data preparation and enhance the platform's data handling capabilities.
- **Improve model interpretability:** Enhance model interpretability by refining feature selection methodologies, enabling users to better understand the factors influencing model predictions.
- **Optimise model performance:** Explore fine-tuning techniques for model parameters to improve both accuracy and efficiency.
- **Incorporate diverse performance metrics:** Introduce a range of performance metrics tailored to meet the needs of various business tasks.

To conclude, the PoC serves as a valuable reference for insurance companies looking to establish collaborative partnerships with cross-sector data providers utilising FL. This approach holds significant potential for leveraging alternative data sources, enhancing risk assessment, improving customer insights, and fostering innovative product development. By accessing diverse datasets, insurers can enrich their understanding of customer behaviour and market trends, driving better decision-making and leading to competitive advantage. However, it is imperative for insurers to conduct a thorough evaluation of its benefits and drawbacks to ensure successful implementation and alignment with organisational goals, as the FL is still in its infancy and has limitations.

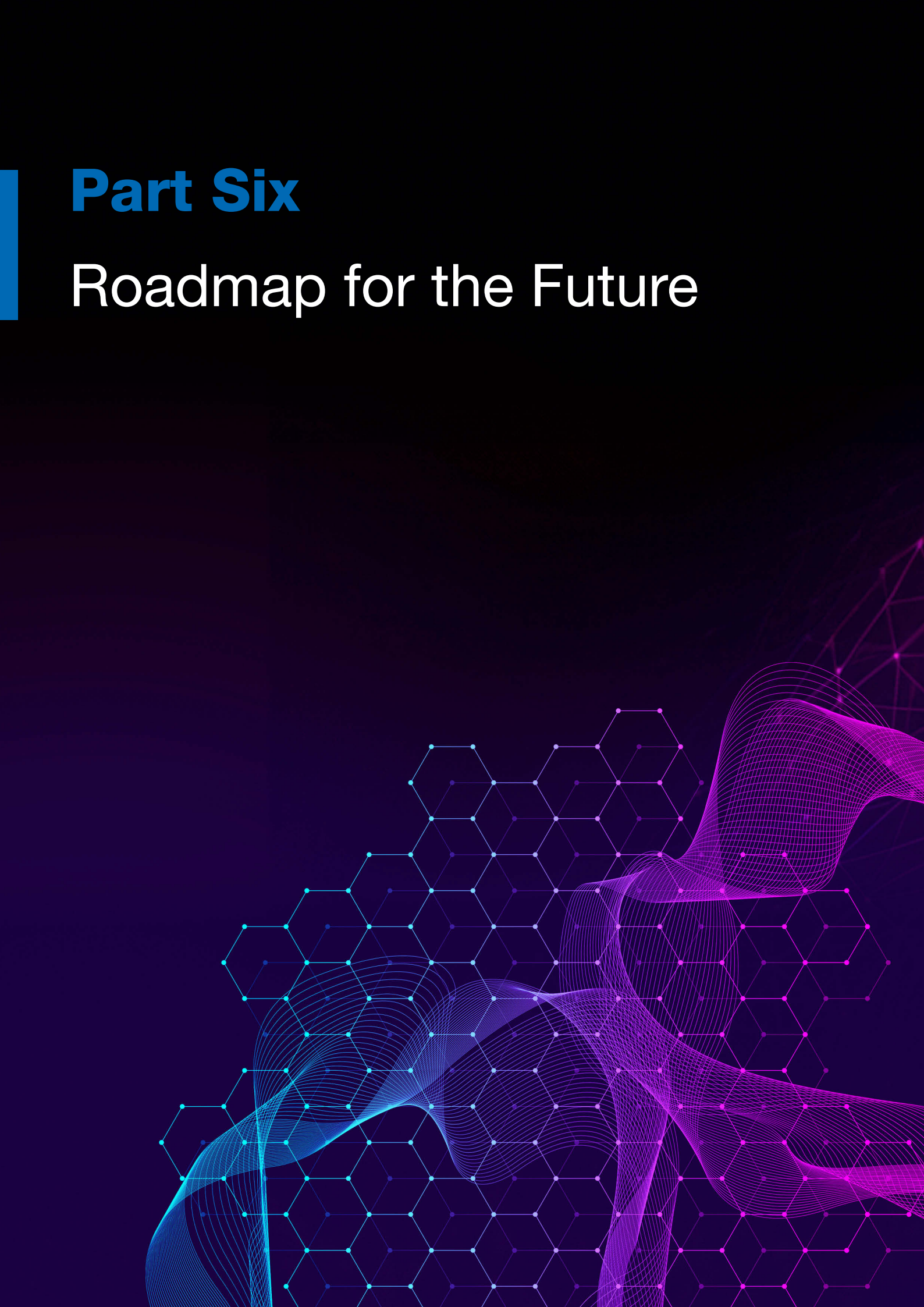
Acknowledgements for contributions to this section:

Company	Contributions
Bowtie & JP Health Bowtie Life Insurance Company Limited FWD Life Insurance Company (Bermuda) Limited Company in Retail Sector (Anonymous) Nova Credit Limited[82] OneDegree Hong Kong Limited	Information and advice on the PoC Work

82 Nova Credit Limited, a Hong Kong-based credit reference agency, decided to cease operations and exit the Credit Data Smart programme in July 2024 due to a depletion of funds.

Part Six

Roadmap for the Future



Part Six:
Roadmap for the Future

This section presents a high-level roadmap with possible future directions for implementing FL in Hong Kong's insurance industry. It outlines strategic priorities in three key areas: the technical landscape, organisational considerations, and the enabling ecosystem. The roadmap aims to help developers create robust and scalable FL solutions, assist organisations in their effective implementation, and support stakeholders in enhancing the digital ecosystem within the insurance sector.

6.1 Technical Roadmap – Advancements in FL Technology

• Optimising the efficiency of FL

As an FL system grows in scale, to involve hundreds or thousands of clients, its efficiency needs to be optimised for greater scalability, responsiveness, and cost-effectiveness. The availability of high-performance devices has lowered the barriers to large-scale FL deployment. Key strategies for improving FL efficiency include:

1. **Adaptive client selection:** Use dynamic algorithms to prioritise which clients participate, based on their computational capabilities, network conditions, and data quality. Techniques such as reinforcement learning or energy aware multi-armed bandit approach can help in selecting the most suitable clients for the FL training round.

2. **Communication-efficient protocols:** Develop communication-efficient protocols that minimise the transmission size of model updates and gradients. Techniques such as gradient compression, pruning (e.g. sparse model representation, model pruning), and quantisation (e.g. weight quantisation, gradient quantisation) can minimise transmission size, boosting communication efficiency.

3. **Flexible and decentralised training:** Develop training methods that allow each participant to work at their own pace, without having to stay perfectly in sync with others, and investigate decentralised architectures to eliminate the need for a central coordinator, reducing communication bottlenecks and improving scalability.

4. **Hardware-software co-design:** Collaborate with hardware vendors to develop specialised FL-optimised hardware accelerators, like edge devices and mobile chipsets. Integrating such co-designed hardware-software solutions can further optimise FL system performance.

• Improving data privacy and security

FL faces challenges from evolving vulnerabilities and potential attacks. To ensure data privacy and security, organisations should consider the following strategies:

1. **Continuous threat monitoring:** Establish robust mechanisms for continuous monitoring and analysis of emerging threats and vulnerabilities in FL systems, focusing on the latest trends, attack vectors, and potential risks. This could include implementing anomaly detection algorithms that identify unusual patterns in model updates or client behaviour.

2. **Adopting industry-wide security standards:** Adopt industry-wide security standards in Hong Kong that align with local regulations and international best practices. Collaborating with regulatory bodies like the HKMA and PCPD can help develop secure frameworks for cross-sector data sharing through FL. Before implementing FL, system developers and users should ensure that the platform has obtained relevant security certifications like ISO/IEC 27001 and ePrivacyseal[83], which demonstrate a commitment to data protection and raise stakeholder trust in FL technology. They should also reference the 3652.1-2020 – IEEE Guide for Architectural Framework and Application of Federated Machine Learning[84] and 2986-2023 – IEEE Recommended Practice for Privacy and Security for Federated Machine Learning[85] when developing their FL applications to ensure compliance with privacy, security, and regulatory requirements.

3. **Encouraging cross-organisation collaboration:** Foster collaboration and knowledge-sharing among organisations to enhance security measures in FL systems. Establishing a consortium of researchers, cybersecurity experts, and industry stakeholders can improve information exchange and facilitate collaborative research. In Hong Kong, partnerships with organisations like Hong Kong Cyberport and ASTRI can drive innovation and support joint initiatives to enhance FL system security.

4. **Establishing security-focused research initiatives:** Develop research initiatives focused on security to address specific vulnerabilities in FL systems. Connecting Hong Kong's academic institutions with industry players can leverage their expertise in cybersecurity research to enhance system security.

5. **Creating testbeds for validation:** Set up testbeds to validate privacy-preserving techniques in a controlled environment. In Hong Kong, these testbeds can be established through collaborations with research institutions and tech hubs like Cyberport, providing a space for experimentation and innovation in security practices for FL systems. Additionally, the Insurtech Sandbox launched by the IA in 2017 is a valuable platform for testing insurance solutions.

6. **Adaptable and scalable defence mechanisms:** Design resilient and adaptable FL architectures in Hong Kong through collaborations between local universities and industry, including developing protocols that incorporate local threat intelligence.

6.2 Organisation Roadmap – Promote FL Adoption

The Organisation Roadmap seeks to create an environment conducive to the industry adoption of FL, with a supportive regulatory framework, industry-wide standards and guidelines, and comprehensive educational programmes to build awareness and skills. It aims to foster collaboration among regulatory bodies, market participants and cross sector stakeholders, develop secure data-sharing infrastructure, and encourage research partnerships. The Organisation Roadmap contains the following key actions:

1. **Education about and promotion of FL:** Collaborate with industry associations (e.g. the Hong Kong Federation of Insurers (HKFI) on educational programmes to raise awareness of FL among insurance professionals. Through workshops, seminars, and online resources, highlight its potential benefits in risk assessment, customer insights, and operational efficiencies, as well as the associated risks.

83 ePrivacyseal awards a data protection seal after conducting an in-depth audit of online and mobile products based on GDPR. It is designed for companies with no direct data processing operations, such as cloud services and SaaS. The certification criteria are continuously updated to ensure compliance with data protection laws. In Hong Kong, the Openhive Federated Learning Platform is the first enterprise-grade federated learning data network to obtain this certification.

84 3652.1-2020 – IEEE Guide for Architectural Framework and Application of Federated Machine Learning provides a blueprint for data usage and model building across organisations while meeting applicable privacy, security and regulatory requirements. It defines the architectural framework and application guidelines for federated machine learning. This guide was published on 19 March 2021.

85 2986-2023 – IEEE Recommended Practice for Privacy and Security for Federated Machine Learning provides recommended practices related to privacy and security for FML, including security and privacy principles, defense mechanisms against non-malicious failures and examples of adversarial attacks on a FML system. This document also defines an assessment framework to determine the effectiveness of a given defense mechanism under various settings. This document was published on 26 April 2024.

2. **Infrastructure development:** Develop a federated data exchange infrastructure modelled on the Commercial Data Interchange (CDI)[86] launched by the HKMA in 2022 to enable secure and seamless data sharing among insurers, to support strong data governance and access control and ensure compliance with relevant regulations, thereby strengthening collaboration in product development and risk management.

3. **Talent development and staff training:** Establish partnerships with local universities and professional training institutions, such as the Vocational Training Council (VTC), to build a skilled workforce proficient in FL methodologies. Such partnerships will result in specialised training programmes and industry conferences for insurance professionals in Hong Kong, covering the technical and operational aspects of FL and equipping staff with the skills to leverage FL technologies effectively.

4. **Partnerships with research institutions and fintech companies:** Establish strategic partnerships with local research institutions and fintech companies such as ASTRI to drive innovation in FL applications, facilitating the development of tailored solutions for the insurance market, and promoting technology integration and scalability.

5. **Establishment of industry-wide standards:** Actively develop regulatory guidelines and industry standards for implementing innovative AI technologies such as FL in the insurance sector. By contributing diverse perspectives and insights, organisations can help ensure compliance, safeguard data security, and promote best practices to build stakeholder trust.

6.3 Ecosystem Roadmap – Cross-sector Collaboration

The Ecosystem Roadmap aims to facilitate cross-sector collaboration, creating synergies that enhance the effectiveness of the insurance sector while contributing to the broader digital transformation of Hong Kong's economy.

1. **Promote cross-sector collaboration:** Encourage partnerships with sectors such as healthcare, banking, research, and startups, to leverage diverse data sources and expertise. For example, collaborations with healthcare providers such as the Hospital Authority can enhance risk assessment by providing access to anonymised health data. Collaborations with government bodies such as the HKMA can facilitate regulatory frameworks that support data sharing and innovation.

2. **Encourage stakeholder engagement:** Engage customers, policymakers, and industry associations, including the HKIA and HKFI, to incorporate their diverse perspectives into FL initiatives. Consultations, forums, and workshops will help stakeholder needs be understood and address data privacy and security concerns, fostering trust and collaboration across the ecosystem.

3. **Leverage existing digital infrastructure:** The insurance sector could potentially leverage existing data sharing infrastructure, such as the HKMA's CDI, to minimise the time and technical resources required for data exchange. In August 2024, the HKMA and the Digital Policy Office (DPO) jointly announced the full operation of CDI and the Government's Consented Data Exchange (CDEG), which facilitates data exchange between the government and banks. Banks can now directly obtain company particulars such as registered addresses or company names to streamline various processes such as fraud detection. The Companies Registry (CR) has become the first party to connect to CDI through the CDEG.

In short, the FL roadmap is a multifaceted strategy to drive continued technical advancements, promote industry adoption, and foster cross-sector collaboration. Key technical focuses include creating efficient, scalable, and secure FL solutions through adaptive client selection, communication-optimised protocols, decentralised training architectures, and specialised hardware. To drive adoption, the roadmap calls for organisational changes to align incentives, modernise data governance, and build internal capabilities. It also emphasises the importance of cross-industry cooperation to enhance data availability, establish common standards, and address emerging privacy and security threats.

86 Commercial Data Interchange (CDI) is a consent-based financial data infrastructure launched by the HKMA to enhance data sharing among financial institutions. It enables the retrieval of commercial data, especially from small and medium-sized enterprises (SMEs), from public and private data providers. CDI supports innovative financial applications like Know-Your-Customer (KYC), credit assessment, and risk management, promoting secure and seamless data exchange in Hong Kong.

Annex A: POC Evaluation

1. Platform System Requirements

The successful deployment of the FL platform requires specific technical infrastructure requirements to be met, for both hardware and software components. Table 35 outlines the minimum hardware and software system requirements for installing and running the platform. For hardware, the requirements include Kubernetes clusters, container registry, managed PostgreSQL database, and ingress controller. For software, the build machine needs Docker and Helm Chart, while the client machine should have the latest version of the Chrome browser. These are needed to ensure that the platform's training performance, including its training speed, is acceptable. The items highlighted in bold in the table could affect performance, scalability, and reliability. Alternative setups will require thorough evaluation to ensure they meet the technical specifications, scalability needs, and compatibility with existing systems, and the expertise of the team available.

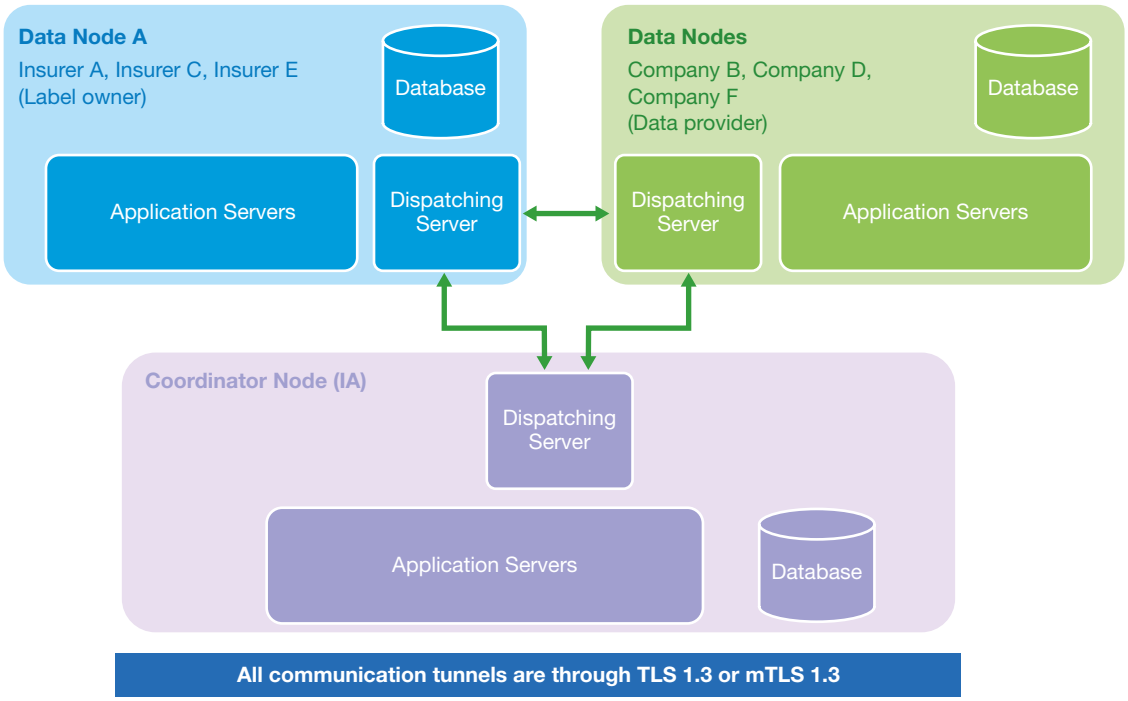
Table 35 Requirements for platform system deployment

Requirement		Details
Hardware Requirements	Kubernetes Service	• 2 Kubernetes clusters, 1 node pool per cluster, and 1 node per node pool • Kubernetes version: 1.22 or above • Node Operating System: Linux • CPU: 8 cores with 3GHz or above • Ram: 32 GB memory or above
	Container Registry	• 100GB or above storage
	Database	• Fully managed database for PostgreSQL • PostgreSQL, version 13 • Performance configuration: Basic[87], 2 vCore(s)[88], 1TB[89] or above
	Application Gateway/Load Balancer/Nginx (for ingress controller)[90]	/
Software Requirements	Build Machine	• Docker – version 20.10 or above • Helm Chart – version 3
	Client Machine	• Chrome Browser – latest version

87 Basic refers to an entry-level performance tier provided by cloud platforms, offering essential resources at a lower cost for moderate workloads.
88 2 vCores indicates a modest compute allocation suitable for small- to medium-scale applications, often provisioned on virtualised infrastructure managed by the cloud provider.
89 The 1TB figure denotes the allocated persistent storage capacity for data, indexes, and logs. The memory is generally tied to the vCore configuration.
90 By default, the platform uses a Kubernetes-native Ingress controller (such as NGINX) or the load balancing services provided by the cloud provider, with configurations based on environment defaults. If a dedicated hardware load balancer is required, additional evaluation of network interface specifications and bandwidth requirements should be conducted.

Figure 17 shows the basic architecture of the FL platform, illustrating the data flow between Data Node A (the Insurer), Data Node B (the Data Provider), and the Coordinator Node (IA).

Figure 17 Platform architecture overview



2. Performance Evaluation Methodology

For the use cases, the following performance metrics were used to evaluate the federated model results against the local model results in terms of effectiveness and capability:

- **Area under the curve (AUC):** AUC measures the ability of a model to distinguish between positive and negative classes. Values range from 0 to 1, with a value of 0.5 indicating no discrimination (e.g. random guessing), and a value of 1 signifying perfect discrimination. In practice, an AUC value above 0.7 is generally considered acceptable, while values above 0.8 are often viewed as strong indicators of good model performance.

- **Gini Index (GI):** The Gini Index quantifies inequality among values. It ranges from 0 (the worst performance) to 1 (the best performance). A higher GI Index score indicates a better classifier performance, while a lower value suggests poorer performance.

- **KS (Kolmogorov-Smirnov) Index:** The KS Index measures the maximum difference between the cumulative distributions of predicted probabilities for positive and negative classes. A higher KS value indicates the model's stronger discriminatory power, with values above 0.3 generally considered indicative of good model performance.

- **Mean Squared Error (MSE):** MSE quantifies the average squared difference between predicted and actual values. Lower values indicate better predictive accuracy, with a value of 0 representing perfect predictions.

- **Ratio of improvement of Federated Learning (RIFL):** RIFL measures how much better the FL model performs by comparison with the local model. A RIFL value of greater than 1 indicates an improvement achieved through FL, while a value of less than 1 suggests no improvement.

3. Details of Experimental Results

3.1 Use Case 1

- **Key performance metrics of local and federated models**

Performance metric comparisons between local and federated models in training and back testing across varied algorithms (Logistic Regression, Boosting, and Neural Network) are shown in Table 36.

Local model results are obtained by training each participant's model using the same algorithm (e.g. Logistic Regression) and the same federated-trained parameters, but exclusively on its own local dataset. These results are then compared with the federated model's performance in back testing. These results help assess the effectiveness of the federated trained model on the local dataset and its potential for generalisation to unseen data.

a. Moderate performance in local setting

In the local setting, all algorithms delivered moderate predictive performance. Logistic Regression showed the highest AUC, the largest KS index, and the lowest MSE, suggesting it captured linear relationships in the data and discriminates between false positives and true positives most effectively. Boosting performed worst, with the highest MSE at 0.0369 and the lowest AUC at 0.7033. However, Neural Network had a worse KS Index than Boosting, suggesting that it struggles in distinguishing between positive and negative classes.

b. Improved model metrics in FL

Comparing the performance metrics, such as AUC values in bold blue and KS Index in bold black, across all models in both local and federated settings reveals that FL generally enhanced model performance. Logistic Regression showed a slight improvement in AUC and Gini Index in the federated setting, while Boosting demonstrated the most significant gains, particularly in AUC, KS Index, and Gini Index, indicating its superior ability to leverage diverse data. Neural Network also showed marginal improvements in the federated setting, with a notable increase in the KS Index. KS Index values, highlighted in bold black, for three algorithms demonstrate an acceptable level of above 0.3.

c. FL with Boosting

Boosting shows the most significant improvement in the federated setting, probably due to the fact that the alternative data provided rich, diverse information having complex interactions with features from the data consumer. This improvement could stem from the existence of a non-linear relationship between the features sourced from Company B and Insurer A, which Boosting can handle effectively.

Table 36 Key performance metrics of local and federated models with different algorithms (Use Case 1)

Model Types	Local				Federated			
	MSE	AUC	KS Index	Gini	MSE	AUC	KS Index	Gini
Logistic Regression	0.0319	**0.7469**	**0.4848**	0.4937	0.0341	**0.7686**	**0.4770**	0.5372
Boosting	0.0369	**0.7033**	**0.4837**	0.4066	0.0366	**0.8046**	**0.6304**	0.6092
Neural Network	0.0360	**0.7381**	**0.3953**	0.4763	0.0355	**0.7458**	**0.4899**	0.4916

Table 37 Ratio of improvement of FL for use case 1

Model Types	GI-based RIFL	MSE-based RIFL
Logistic Regression	ρ_{LR} improvement = 8.80%	α_{LR} improvement = 0%
Boosting	ρ_{Boost} improvement = 49.83%	α_{Boost} improvement = 0.83%
Neural Network	ρ_{NN} improvement = 3.21%	α_{NN} improvement = 1.38%

• Evaluation of the ratio of improvement of FL

The evaluation results on the ratio of improvement of FL, namely GI-based RIFL and MSE-based RIFL, are summarised in Table 37.

i. FL improvement with Boosting

The majority of the models were improved to varying degrees by applying FL techniques. Boosting demonstrated the most significant improvements in terms of GI-based RIFL, with a value of ρ_{Boost} improvement = 49.83%. This indicates a substantial relative improvement in Boosting's performance under the FL approach.

ii. FL improvement with Logistic Regression

For Logistic Regression, the RIFL score was zero when evaluated by the Mean Squared Error (MSE), indicating that the federated model performed worse than the locally trained model. Consequently, the RIFL was capped at zero, which is not useful. This issue may have stemmed from an imbalance in the data in which some outcomes were more common than others, making it difficult for the model to accurately predict continuous values.

iii. FL improvement with Neural Network

Neural Network experienced positive improvements in both the GI-based RIFL and MSE-based RIFL. The GI-based RIFL for Neural Network was 3.21%, while the MSE-based RIFL was 1.38%.

3.2 Use Case 2

The target of the model was to predict the probability of insurance claims. The incorporation of alternative data, specifically historical health data, had a significant impact on the performance of the predictive models.

• Key performance metrics of local and federated models

Table 38 provides a detailed comparison of key performance metrics for both local and federated models using different algorithms.

a. **Boosting outperformed in both local and federated settings, excelling in all metrics (MSE, AUC, KS Index, Gini).** It specifically showed the lowest MSE, and the highest scores in AUC, Gini Index, and KS Index. In the context of insurance claim prediction, where data is typically skewed with fewer claim instances than non-claims, Boosting proved to be efficient. Because it concentrated more on incorrectly classified instances, it enhanced the model's performance on the less represented class in this use case, namely claim instances.

Table 38 Key performance metrics of local and federated models with different algorithms (Use Case 2)

Model Types	Local				Federated			
	MSE	AUC	KS Index	Gini	MSE	AUC	KS Index	Gini
Logistic Regression	0.2504	**0.5308**	**0.0783**	0.0617	0.2306	**0.6534**	**0.2319**	0.3067
Boosting	0.2621	**0.5471**	**0.1457**	0.0942	0.2081	**0.7945**	**0.4394**	0.5890
Neural Network	0.2499	**0.5104**	**0.0434**	0.0207	0.2452	**0.5351**	**0.1125**	0.0701

b. **Federated models outperformed local models across all three metrics.** The enhanced performance of federated models is likely attributable to their training process, which combined traditional insurance data from Insurer C with historical health data from Company D. The resulting collaborative dataset was larger and more diverse than the local dataset, enhancing the model's performance.

c. **AUC and Gini Index demonstrated the federated model's superior predictive accuracy.** The AUC score for the local model ranged from 0.5104 to 0.5471, indicating performance only slightly better than random guessing. By contrast, the federated model achieved a generally higher AUC range of 0.5351 to 0.7945, suggesting improved predictive capability. Notably, the Gini Index revealed a more pronounced performance gain in the federated model, highlighting its superior ability to distinguish between classes compared to the local model.

d. **Among the three algorithms evaluated, Neural Networks performed the worst.** In this use case, where there is a strong correlation between the prediction label and training features, simpler models like Logistic Regression and Boosting can adequately capture and explain the relationship. By contrast, Neural Networks typically excel in scenarios with weaker correlations or more complex, non-linear data structures.

• Evaluation of the ratio of improvement of FL

Table 39 summarises the evaluation results, specifically the GI-based RIFL and MSE-based RIFL.

Table 39 Ratio of improvement of FL for use case 2

Model Types	GI-based RIFL	MSE-based RIFL
Logistic Regression	ρ_{LR} improvement > 3 times	α_{LR} improvement = 7.90%
Boosting	ρ_{Boost} improvement > 5 times	α_{Boost} improvement = 20.60%
Neural Network	ρ_{NN} improvement > 2 times	α_{NN} improvement = 1.91%

i. Improvements in RIFL with FL

All models exhibited positive improvements in AUC-based RIFL and MSE-based RIFL, demonstrating the effectiveness of FL in enhancing model performance. The GI-based RIFL for Logistic Regression achieved a remarkable increase of over 3 times that of the local model, while the MSE-based RIFL showed a notable improvement of 7.90%. Boosting experienced the largest gains, with the GI-based RIFL increasing by more than 5 times that of the local model, and the MSE-based RIFL showing a substantial improvement of 20.60%.

ii. Higher GI-Based RIFL Compared to MSE-Based RIFL

The results demonstrate the substantial performance improvements in GI-based RIFL for all three ML models compared to their individual performances on separate local datasets. The GI-based RIFL achieved a striking increase of over five times, a level not observed in the MSE-based RIFL.

This can be attributed to the relationship between the Gini Index and AUC. If the local model's AUC is 0.5 (indicating randomness), even a small improvement in AUC can lead to a significant increase in the Gini Index.

Simulation results

To simulate and validate the scenario that showed an extreme improvement in the Gini Index (GI-based RIFL), an open-source insurance-related dataset from Kaggle[91] was collected. This comprised 2,000 rows and 44 features after feature selection. In the simulation, we assumed that Data Consumer A and Data Provider B are joining the training, and vertically divided the dataset into two parts. Data Consumer A contributed only 1 feature, while Data Provider B contributed 43 features.

Table 40 AUC scores and Gini Index of local and federated models for the simulated experiments

Model Types	Local		Federated	
	AUC	Gini Index	AUC	Gini Index
Logistic Regression	0.5157	**0.0314**	**0.5718**	0.1435
Boosting	0.5010	**0.0020**	**0.6649**	0.3298
Neural Network	0.5157	**0.0314**	**0.6372**	0.2744

91 https://www.kaggle.com/datasets/moneystore/agencyperformance/data

Table 40 presents the simulated results of the AUC scores and Gini Index from both local and federated models. It shows that in the local model, all AUC scores fall within the range of 0.5010 to 0.5157, indicating values close to 0.5. Their corresponding GI-based RIFLs in three algorithms demonstrate quite major improvements: ρ_{LR} improvement > 3 times, ρ_{Boost} improvement > 163 times, and ρ_{NN} improvement > 7 times. These extreme values confirm that when there is an unequal distribution of features among participating parties in FL, or when the local model has an AUC value close to 0.5, the federated model will generate a significant improvement over the local model in terms of the Gini Index.

3.3 Use Case 3

- **Key Performance Metrics of Local and Federated Models**

Table 41 provides a detailed comparison of key performance metrics for both local and federated models using different algorithms. As the data provider did not proceed to Boosting, some results in the table are absent, shown by "/".

Table 41 Key performance metrics of local and federated models with different algorithms (Use Case 3)

Model Types	Local				Federated			
	MSE	AUC	KS Index	Gini	MSE	AUC	KS Index	Gini
Logistic Regression	0.1586	0.9632	0.8828	0.9264	0.1573	0.9296	0.8373	0.8591
Neural Network	0.0311	**0.9409**	0.9927	0.8819	0.0181	**0.9874**	0.9661	0.9749
Boosting	0.0373	0.9241	0.8481	0.8481	/			

a. Neural Network leads in local setting

In the local setting, all algorithms achieved satisfactory KS Index values above 0.3 and AUC scores ranging from 0.9241 to 0.9632, with Logistic Regression demonstrating the best performance. However, Neural Network excelled in distinguishing positive and negative outcomes, evidenced by its lower MSE and higher KS Index compared to both Logistic Regression and Boosting.

b. Neural Network leads in federated setting

Neural Network outperformed Logistic Regression in the federated setting, exhibiting lower MSE and higher AUC (in bold blue), KS Index, and Gini Index. Its superior performance may stem from its ability to automatically capture interactions between insurance and credit features, such as identifying higher risk in policyholders with high claim counts and low credit scores. This is unlike Logistic Regression, which requires manual specification of these interactions.

c. Federated model outperforms local model

The local model here is derived from federated learning, which allows it to incorporate additional information from data partners as a form of enhancement. As a result, metrics such as AUC, KS Index and Gini Index may exceed those of the true local model trained solely by the data consumer, or even those of the federated model. However, the federated model always has a lower MSE than the local model once it has achieved optimal performance.

d. Incompleteness of federated model

Boosting demonstrated satisfactory local results, whereas the federated model remained incomplete due to the exit of the data provider.

• Evaluation of the Ratio of Improvement of FL

The evaluation results, namely GI-based RIFL and MSE-based RIFL, are summarised in Table 42.

Table 42 Ratio of improvement of FL for use case 3

Model Types	GI-based RIFL	MSE-based RIFL
Logistic Regression	ρ_{LR} improvement = 0%	ρ_{LR} improvement = 0.82%
Neural Network	ρ_{NN} improvement = 10.55%	α_{NN} improvement = 41.97%
Boosting	/	/

i. The results for Logistic Regression were unexpected in terms of Gini Index (or AUC score), with the incorporation of alternative data having a negative effect on the federated model.

ii. The GI-based RIFL for the Logistic Regression model was ρ_{LR} improvement = 0%, a lower AUC value than the local model. Therefore, GI-based RIFL was capped to 0 to show its unavailability. Conversely, the MSE-based RIFL improvement for Logistic Regression was 0.82%, meaning the federated model had a 0.82% lower MSE than the local model.

iii. Notably, Neural Network showed improvements in both the GI-based RIFL and MSE-based RIFL. The GI-based RIFL improvement for Neural Network was 10.55%, while the MSE-based RIFL improvement was 41.97%.

Simulation results

To replicate the scenario of a negative GI-based RIFL in Logistic Regression, open-source insurance-related dataset from Kaggle[92], consisting of 2,000 rows and 53 features, was again utilised. On the assumption that Data Consumer A and Data Provider B were joining the training, the dataset was vertically split into two parts, with Data Consumer A contributing 14 features and Data Provider B 39 features. Certain features have missing entries.

The three algorithms, Logistic Regression, Boosting, and Neural Network, were implemented using two distinct methods to handle the missing entries. The first method involves filling the missing year with a large number, like 999,999. Another method leverages the one-hot encoding technique, which encodes the missing entry with 0 and 1, where 0 represents the absence of the entry and 1 its presence.

Table 43 shows that the AUC scores for Logistic Regression, Boosting, and Neural Network in both local and federated models range from 0.8993 to 0.9731, indicating satisfactory performance. However, when missing values were filled with large numbers, the federated Logistic Regression model had a slightly lower AUC than the local model, while Boosting and Neural Network performed better in the federated setting.

Table 43 Local and federated models in different models with two missing value handling methods

Model Types	Missing value handling	Local				Federated			
		MSE	AUC	KS Index	Gini	MSE	AUC	KS Index	Gini
Logistic Regression	**Filling with big number**	0.1400	0.9037	0.6924	0.8074	0.1364	0.9028	0.6872	0.8057
	One-Hot encoding	0.1429	**0.8993**	0.6679	0.7986	0.1271	**0.9038**	0.6776	0.8076
Boosting	**Filling with big number**	0.1202	0.9174	0.6641	0.8348	0.0728	0.9620	0.7924	0.9239
	One-Hot encoding	0.1513	0.8748	0.6409	0.7496	0.0583	0.9731	0.8346	0.9463
Neural Network	**Filling with big number**	0.1203	0.9129	0.7001	0.8258	0.1117	0.9154	0.6993	0.8307
	One-Hot encoding	0.1140	0.9098	0.6915	0.8195	0.1112	0.9127	0.6993	0.8253

92 https://www.kaggle.com/datasets/moneystore/agencyperformance/data

Table 44 Ratio of improvement of FL for the simulated experiments

Model Types	Filling with big number	
	GI-based RIFL	MSE-based RIFL
Logistic Regression	ρ_{LR} improvement= 0%	α_{LR} improvement = 2.57%
Boosting	ρ_{Boost} improvement = 10.67%	α_{Boost} improvement = 39.43%
Neural Network	ρ_{NN} improvement = 0.59%	α_{NN} improvement = 7.15%
Model Types	One-Hot encoding	
	GI-based RIFL	MSE-based RIFL
Logistic Regression	ρ_{LR} improvement = 1.13%	ρ_{LR} improvement = 10.38%
Boosting	ρ_{Boost} improvement = 26.24%	ρ_{Boost} improvement = 61.47%
Neural Network	ρ_{NN} improvement = 0.71%	ρ_{NN} improvement = 2.46%

Table 44 reveals that incorporating features from Data Provider B with Data Consumer A resulted in a GI-based RIFL of ρ_{LR} improvement = 0%, a metric that is not informative or useful. By contrast, using one-hot encoding for missing values improved the federated model's performance, so that it achieved a positive GI-based RIFL of 1.13%.

These results suggest that improperly handled missing values can significantly harm the federated model's performance, particularly for Logistic Regression, which is sensitive to data quality. Unlike Neural Network, which functions as a "black box", Logistic Regression relies heavily on input feature values for training and predictions.

Annex B: List of Acronyms

Acronyms	Full Form	Acronyms	Full Form
APIs	Application programming interfaces	**IoT**	Internet of Things
AUC	Area Under the Curve	**KNN**	K-Nearest Neighbours
BDA	Big data analytics	**KS**	Kolmogorov-Smirnov
CDFL	Cross-device federated learning	**LR**	Logistic Regression
CIMM	Confidential identity matching module	**ML**	Machine learning
CNN	Convolutional Neural Networks	**MLP**	Multi-Layer Perceptron
CSFL	Cross-silo federated learning	**MSE**	Mean squared error
DP	Differential privacy	**NN**	Neural Network
DPPs	Data Protection Principles	**PCPD**	Office of the Privacy Commissioner for Personal Data
EU	European Union	**PETs**	Privacy-enhancing technologies
FATE	Federated AI Technology Enabler	**PI**	Personal information
FL	Federated learning	**PII**	Personally identifiable information
FLUTE	Federated Learning Utilities and Tools for Experimentation	**PIPL**	Personal Information Protection Law
FN	False negative	**PoC**	Proof-of-Concept
FP	False positive	**PSI**	Private set intersection
FPR	False Positive Rate	**RIFL**	Ratio of Improvement of Federated Learning
FTL	Federated transfer learning	**RNN**	Recurrent Neural Networks
FTSM	Fast-training strategy module	**ROC**	Receiver Operating Characteristic
GBA	Greater Bay Area	**SMPC**	Secure multi-party computation
GDPR	General Data Protection Regulation	**TEE**	Trusted execution environments
GI	Gini Index	**TN**	True negative
GINA	Genetic Information Nondiscrimination Act of 2008	**TP**	True positive
HE	Homomorphic encryption	**TPR**	True Positive Rate
HFL	Horizontal federated learning	**UBI**	Usage-based insurance
HKID Card	Hong Kong Identity Card	**VFL**	Vertical federated learning

Annex C: Glossary of Key Terms

Term	Definition
Application programming interface (API)	A set of protocols, routines, and tools that allow different software applications to communicate with each other.
Artificial Intelligence (AI)	Technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.
Big Data Analytics (BDA)	The process of examining large and complex datasets to uncover hidden patterns, correlations and insights, often used to inform decision-making in various sectors, including insurance.
Common Data Interchange (CDI)	A next-generation financial data infrastructure that enables more efficient financial intermediation in the banking system and is enhancing financial inclusion in Hong Kong. The Hong Kong Monetary Authority (HKMA) launched a CDI in October 2022.
Collaboration Platforms	Tools and frameworks that facilitate federated learning and data sharing among data partners.
Confidential Identify Matching Module (CIMM)	A newly developed module which employs a hash function and the homomorphic encryption technique to securely match identities across different clients, and includes a neutral third party to distribute the matched results.
Convolutional Neural Network (CNN)	A specialised deep learning architecture designed for processing structured grid-like data, such as images. CNNs automatically learn spatial hierarchies of features through convolutional operations, making them highly effective for tasks like image recognition and object detection.
Cross-device federated learning (CDFL)	A decentralised machine learning approach where models are trained collaboratively across a large number of devices—such as smartphones or edge devices—without centralising raw data.
Cross-silo Federated Learning (CSFL)	A decentralised machine learning approach where a small number of large, trusted organisations (silos) collaboratively train a shared model without directly sharing their raw data.
Data Anonymisation	The process of removing personally identifiable information from datasets to protect privacy.
Data Governance	Policies and practices that ensure data is managed properly, including data quality, data security, and compliance with regulations.
Data Node	A device or entity that holds local data used for training machine learning models.

Term	Definition
Decentralised Machine Learning	The use of machine learning techniques in a distributed manner, where data processing and model training occur on multiple local devices or nodes without the need for a centralised server.
Deep Learning	An artificial intelligence (AI) method that teaches computers to process data in a way inspired by the human brain. Deep learning models can recognise complex pictures, text, sounds, and other data patterns, and produce accurate insights and predictions.
Differential Privacy (DP)	A mathematical framework for ensuring privacy in datasets by adding controlled noise to data or query results. It guarantees that the inclusion or exclusion of any single individual's data does not significantly affect the result of an analysis or query, making it impossible to confidently identify individuals while preserving useful aggregate information.
Data Protection Principles (DPPs)	A set of six core rules that govern how personal data shall be collected, handled, stored, and used by organisations under Hong Kong's Personal Data (Privacy) Ordinance (PDPO).
Edge Nodes	A device located at the periphery of a network, close to data sources, that processes, filters, and analyses data locally.
Encryption	The process of transforming readable plaintext into unreadable ciphertext to mask sensitive information from unauthorised users.
Federated AI Technology Enabler (FATE)	An open-source federated learning framework developed to enable secure, collaborative AI model training across multiple parties without sharing raw data.
Federated Learning (FL)	A machine learning approach that enables multiple participants or devices to collaboratively train a shared model while keeping all the training data decentralised.
False Negative (FN)	An error in statistical testing where a model incorrectly predicts the absence of a condition when it actually exists.
False Positive (FP)	An error in statistical testing where a model incorrectly predicts the presence of a condition when it does not actually exist.
Federated Transfer Learning (FTL)	Enables collaborative model training across different organisations or devices with heterogeneous data, even when their datasets have non-overlapping samples and features.
Fast Training Strategy Module (FTSM)	A newly developed module that enables model updates through matrix manipulation, each participant independently computing their respective portion of the matrix so as to improve training efficiency.
Generative Adversarial Network (GAN)	A type of deep learning model designed to generate synthetic data that closely resembles real data.
General Data Protection Regulation (GDPR)	A comprehensive EU data privacy law that governs how organisations collect, process, store, and share personal data of individuals.

Term	Definition
Global Model	An aggregated model that combines the insights learned from multiple local training models across various devices.
Gini Index (GI)	Measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.
Gross Premiums	In relation to a financial year of an insurer: a) premiums after deduction of discounts specified in policies, or refunds of premiums made in respect of any termination or reduction of risks, but before deduction of premiums for reinsurance ceded and of commissions payable by the insurer; and b) premiums receivable by the insurer under reinsurance contracts accepted by the insurer.
Hash	A fixed-size string of characters generated by a cryptographic algorithm, representing data in a unique format, commonly used in DLT networks for data integrity.
Homomorphic Encryption (HE)	A form of encryption that allows computations to be performed on encrypted data without first having to decrypt it.
Horizontal Federated Learning (HFL)	A decentralised machine learning approach where participants share the same feature space but have different data samples.
Internet of Things (IoT)	Refers to the network of physical objects or “things” embedded with sensors, software, and other technologies for the purpose of collecting data and exchanging it with other devices and systems over the internet.
Interoperability	The ability of applications and systems to securely and automatically exchange data irrespective of geographical, political, or organisational boundaries.
K-Nearest Neighbours (KNN)	A non-parametric method that classifies a new case based on how its neighbours are classified.
Kolmogorov-Smirnov (KS) Index	Measures the maximum difference between the cumulative distributions of predicted probabilities for positive and negative classes.
Large Language Model (LLM)	A type of machine learning model designed for natural language processing tasks such as language generation.
Local Training Model	A machine learning model that is trained on data residing on a specific client device (data node) without that data being shared with a central server.
Logistic Regression (LR)	A supervised machine learning algorithm used for binary classification.
Machine Learning (ML)	A subset of AI that focuses on developing algorithms that enable computers to learn from and make predictions based on data. without being explicitly programmed.
Machine learning Operations (MLOps) are a set of practices	A set of practices that automate and simplify machine learning workflows and deployments.

Term	Definition
Multi-Layer Perceptron	A class of feedforward artificial neural network composed of multiple layers of interconnected neurons.
Mean Squared Error (MSE)	Measures the average squared difference between estimated values and the true value.
Model Updates	Changes or adjustments made to a machine learning model.
Natural Language Processing (NLP)	A subfield of computer science and artificial intelligence (AI) that uses machine learning to enable computers to understand and communicate using human language.
Neural Network (NN)	A machine learning programme, or model, that makes decisions in a manner similar to the human brain, by using processes that mimic the way biological neurons work together to identify phenomena, weigh options and arrive at conclusions.
Optical Character Recognition (OCR)	The process that converts an image of a text into a machine-readable text format.
Personal Data (Privacy) Ordinance (Cap. 486) (PDPO)	An Ordinance in Hong Kong that protects the privacy of individuals in relation to personal data, and that provides for matters incidental to or connected with data privacy.
Personally Identifiable Information (PII)	Any information connected to a specific individual that can be used to uncover that individual's identity, such as HKID card number, full name, email address or phone number.
Privacy-enhancing Technologies (PETs)	Technologies, tools, techniques, and practices designed to protect the privacy of individuals.
Proof-of-Concept (PoC)	Evidence, typically deriving from an experiment or pilot project, which demonstrates that a design concept, business proposal, etc. is feasible.
Private Set Intersection (PSI)	A secure multiparty computation cryptographic technique that allows two parties holding data sets to compare encrypted versions of these sets and compute their intersection.
Role-based Access Controls	A security model that restricts system access based on user roles rather than individual identities.
Ratio of Improvement of Federated Learning	Measures how much better a FL approach performs in comparison to a local model.
Recurrent Neural Network (RNN)	A type of artificial neural network designed for sequential data, such as speech.
Receiver Operating Characteristic (ROC) curve	A graphical plot that illustrates the performance of a classifier model at varying threshold values.

Term	Definition
Salt	A random value added to input data (e.g. passwords) before it is processed by a hashing algorithm. This ensures that even if two inputs are the same, their hashed outputs will be different.
Shapley Additive Explanations (SHAP)	A method used in machine learning for explaining the output of a machine learning model. It is based on Shapley values from cooperative game theory and provides a unified approach to interpret the predictions of a wide variety of models, including complex deep learning models.
Secure Multi-party Computation (SMPC)	A subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private.
Stochastic Gradient Descent (SGD)	An iterative method for optimising an objective function with suitable smoothness properties (e.g. differentiable).
Trusted Execution Environment (TEE)	An environment for executing code in a secure area of a processor.
TensorFlow Federated	An open-source framework for machine learning and other computations on decentralised data.
True Negative (TN)	An indicator in statistical testing where a model correctly predicts the negative class.
True Positive (TP)	An indicator in statistical testing where a model correctly predicts the positive class.
True Positive Rate (TPR)	Measures a model's ability to correctly identify positive cases out of all actual positives.
Usage-based Insurance (UBI)	A type of auto insurance that calculates premiums based on real-time driving behaviour, rather than traditional factors like age or credit score.
Voluntary Health Insurance Scheme (VHIS)	A policy initiative implemented by the Hong Kong Health Bureau to regulate indemnity hospital insurance plans provided by insurance companies to individuals. Participation in the scheme is voluntary for both insurance companies and consumers.
Vertical Federated Learning (VFL)	A privacy-preserving machine learning paradigm where different parties hold different features of the same set of samples.
Explainable Artificial Intelligence (XAI)	A set of processes and methods that allows human users to comprehend and trust the results and outputs created by machine learning algorithms.


25th
TECH FOR EVERYONE
ASTRI
Hong Kong Applied Science and Technology Research Institute
香港應用科技研究院


保險業監管局
Insurance Authority