

Module Outline

Professional Qualification:	Certified Fintech Professional (Specialist - AI and Big Data Stream) (CPFinT(S-AIBD)) (QF Level 6)
Programme Title:	Postgraduate Certificate for ECF on Fintech (Specialist – AI and Big Data Stream) - Completion of M5 & M6
Module Title:	Artificial Intelligence and Big Data (M6)*
Credit:	30 (21 contact hours, 3 examination hours and about 276 self-study hours)
Teaching/Training Activities:	Training Class
Pre-requisite:	<ul style="list-style-type: none">• Professional Certificate for ECF on Fintech awarded by HKIB; OR• Grandfathered for ECF on Fintech (Core Level) by HKIB• Grandfathered for at least one stream or track under ECF on Fintech (Professional Level) by HKIB
Remarks:	* Pitched at QF Level 6

Professional Qualification Objective

The Professional Qualification of Specialist Track is developed to nurture a sustainable talent pool of more experienced and specialised expert level Fintech practitioners for the banking industry. Learners will achieve mastery within a specific Fintech domain on technology principles and applications in the banking industry to advise on Fintech use cases, solution architecture, and implementation management.

Professional Qualification Intended Outcomes (PQIOs) for Specialist Track (M5, M6, M7, M8, M11)

Upon completion of the Professional Qualification, candidates should be able to:

PQIO 1:	Drive practical Fintech adoption in a range of applicable business contexts.
PQIO 2:	Acquire in-depth knowledge and design rationale of Fintech solutions

	and applications in banks.
PQIO 3:	Evaluate new technologies and vendors in the market and supervise the application of technology concepts of relevant Fintech disciplines in solution design for addressing specific business scenarios and user pain points.
PQIO 4:	Formulate policies for the design, development, and implementation of Fintech solutions.
PQIO 5:	Formulate the implementation approach and guidelines in Fintech projects after reviewing the solution architecture
PQIO 6:	Establish and enforce policies and standards on Fintech software engineering across different coding practices, design patterns, production processes, and system operations.

Module Objective

The module aims to identify current technology trends, opportunities, challenges complexities and risks in developing Big Data (BD) & Artificial Intelligence (AI) projects and recognize the basic principles and good practice for Responsible AI.

Module Intended Outcomes (MIOs) & Units of Competencies (UoCs)

Upon completion of the Module 6, candidates should be able to:

MIOs	Intended Outcome / Competence	Unit of Competencies (UoCs)
MIO 1:	Demonstrate proficiency in the advanced technical concepts, principles, and practical demonstrations of the emerging Fintech topics to execute problem framing and strategy analytics in conducting Fintech solution development.	109398L5 / 107431L5 109400L5 / 107432L5
MIO 2:	Interpret and leverage current and emerging Fintech use cases, project management techniques, and solution implementation, operation, and maintenance methodologies.	109395L6 109577L6 / 107590L6
MIO 3:	Recognise the use and applications of Artificial Intelligence and Big Data leveraging cloud services related to data, and model.	109376L6 109395L6

		109400L5 / 107432L5 109577L6 / 107590L6
--	--	--

Assessment Activity

Type of Assessment Activity	PQIO/MIO	Weighting (%)
Examination	MIO 1 - 3	100%

Examination Format and Duration

Paper 1

Time allowed: 1.5 hours

Format: Multiple choice (closed-book; 50 questions - each 1 mark)

Paper 2

Time allowed: 1.5 hours

Format: Essay Type Questions (open-book; choose 2 out of 3 - each 25 marks)

Pass mark: 50% for combining both papers, failed candidate (<50% as a whole) needs to retake both papers in future attempt.

CPFinT(S-AIBD) Certification Requirement

- Completion of Module 5 and Module 6 of the ECF on Fintech (Professional Level) training; and
- Pass examination of both Modules; and
- Having at least 3 year of relevant work experience in Fintech projects and / or any of the Professional Level functions in Fintech Solution Development, Fintech Product Design, Fintech Strategy and Management or Regtech; and
- Currently as a Relevant Practitioner

Syllabus

Chapter 1: Data Analytics, Artificial Intelligence and Machine Learning
--

1	<ul style="list-style-type: none"> - Overview of data analytics <ul style="list-style-type: none"> • Data analytics use cases in business • Types of data analytics • Data science lifecycle
2	<ul style="list-style-type: none"> - Overview of Artificial Intelligence (AI) <ul style="list-style-type: none"> • What is artificial intelligence? • Artificial intelligence vs. machine learning • Artificial intelligence vs. data science
3	<ul style="list-style-type: none"> - Overview of Machine Learning (ML) <ul style="list-style-type: none"> • Supervised learning • Unsupervised learning • Reinforcement learning • Machine learning lifecycle
4	<ul style="list-style-type: none"> - AI implementation practices <ul style="list-style-type: none"> • Enterprise Architecture (EA) in general • AI model deployment pattern • Cloud-native design & architecture • MLOps vs. DevOps: unique challenges & solutions in AI implementation • MLOps frameworks and tools (including <i>Data Analytics as a Service</i> and <i>Machine Learning as a Service</i>)

Chapter 2: Data Model and Storage

1	<ul style="list-style-type: none"> - Data sources and format <ul style="list-style-type: none"> • Traditional vs non-traditional data sources • Structured, unstructured and semi-structure data format • The 4Vs of Big Data
2	<ul style="list-style-type: none"> - Relational databases <ul style="list-style-type: none"> • Spreadsheets vs relational databases • Conceptual design of relational databases

3	<ul style="list-style-type: none"> • Data query in relational database systems <ul style="list-style-type: none"> - Data warehouses <ul style="list-style-type: none"> • OLTP vs OLAP • Database warehouse design: multidimensional data model • OLAP cube operations
4	<ul style="list-style-type: none"> - NoSQL databases <ul style="list-style-type: none"> • Types of NoSQL databases • Considerations in choosing SQL or NoSQL
5	<ul style="list-style-type: none"> - Big data storage & computing <ul style="list-style-type: none"> • Hadoop technology stack • HDFS distributed storage • MapReduce distributed processing • YARN resource management
6	<ul style="list-style-type: none"> - Cloud storage <ul style="list-style-type: none"> • Cloud vs On-Premise • Type of cloud storage • Data lake as a multi-purpose storage facility

Chapter 3: Supervised Learning and Applications

1	<ul style="list-style-type: none"> - Overview and applications of supervised learning <ul style="list-style-type: none"> • Supervised learning workflow • Applications of classification • Applications of regression
2	<ul style="list-style-type: none"> - Regression models <ul style="list-style-type: none"> • Popular regression models • Encoding of categorical data • Evaluation metrics of regression
3	<ul style="list-style-type: none"> - Classification models <ul style="list-style-type: none"> • Popular classification models

4	<ul style="list-style-type: none"> • Distance measure & data normalization • Evaluation metrics of classification <p>- Ensemble learning</p> <ul style="list-style-type: none"> • Bagging method and a sample model • Boosting method and a sample model • Stacking method
5	A use case in banking and finance

Chapter 4: Unsupervised Learning and Applications

1	<p>- Overview and applications of unsupervised learning</p> <ul style="list-style-type: none"> • Unsupervised learning workflow • Applications of clustering • Applications of dimensionality reduction
2	<p>- Clustering</p> <ul style="list-style-type: none"> • Popular clustering models • Evaluation metrics of clustering • Cluster profiling
3	<p>- Dimensionality reduction</p> <ul style="list-style-type: none"> • The curse of dimensionality • Popular models of dimensionality reduction
4	<p>- Recommendation systems</p> <ul style="list-style-type: none"> • Recommendation systems in production • Overall architecture of a recommendation system • Popular recommendation models
5	- Use cases in banking and finance

Chapter 5: Deep Learning and Applications

1	- Overview of deep learning <ul style="list-style-type: none"> • Deep learning vs traditional ML • Applications of Computer Vision (CV) • Applications of Natural Language Processing (NLP)
2	- Neural network fundamentals <ul style="list-style-type: none"> • Basic structure and major types of networks • Training a neural network • Hyperparameters in neural networks
3	- Computer vision – convolutional neural network <ul style="list-style-type: none"> • Computer vision tasks • CNN architecture
4	- Natural language processing – recurrent neural network <ul style="list-style-type: none"> • Challenges in NLP • Word embedding • RNN based language model
5	- Chatbot development <ul style="list-style-type: none"> • Selective model • Generative model • Chatbot development framework and tools
6	A use case in banking and finance

Chapter 6: Python Hands-on Lab on Data Analytics

1	- Overview of Python development environment <ul style="list-style-type: none"> • What Python is/isn't for • Installation and setup • Development environment • Essential packages
2	- Basic syntax of Python <ul style="list-style-type: none"> • Operators

<p>3</p> <p>4</p> <p>5</p> <p>6</p>	<ul style="list-style-type: none"> • Built-in data types: tuple, list, set, dictionary • NumPy package and ndarray • Pandas package and dataframe <p>- Case – credit card default prediction by classification</p> <p>- Case – customer segmentation by clustering</p> <p>- Introduction of PyTorch</p> <ul style="list-style-type: none"> • PyTorch installation and setup • Tensors and data import • Building models with PyTorch • Training and testing models with PyTorch <p>- Case – scenery image classification by CNN</p>
<p>Chapter 7: ML Engineering Strategies and MLOps</p>	
<p>1</p> <p>2</p> <p>3</p>	<p>- Risks and concerns of AI</p> <ul style="list-style-type: none"> • Need large volume of training data and labels • Algorithm bias • Data drift and concept drift • Invasion of privacy • Lack of explainability • Vulnerable to adversarial attacks <p>- Technological developments in mitigating AI risks</p> <ul style="list-style-type: none"> • Data augmentation • Continuous monitoring • Active learning and transfer learning • Differential privacy and secure multiparty computation • Federated learning • Global and local interpretability techniques <p>- General guideline of Responsible AI</p> <ul style="list-style-type: none"> • Fairness, reliability, accountability, transparency and interpretability, privacy and

4	<p>security</p> <ul style="list-style-type: none"> - AI governance related regulations <ul style="list-style-type: none"> • Hong Kong: HKMA's Guiding Principles on Consumer Protection Aspects in Respect of the Use of BDAI • Hong Kong: HKMA's High-level Principle on AI • Hong Kong: PCPD's Guidance on Ethical Development and Use of AI • China: Provisions on the Scope of Necessary Personal Information for Common Types of Mobile Internet Applications 《常见类型移动互联网应用程序必要个人信息范围规定》 • China: Administrative Provisions on Algorithm Recommendation of Internet Information Services 《互联网信息服务算法推荐管理规定》 • EU: General Data Protection Regulation (GDPR) and its impact on AI • EU: Ethics Guidelines for Trustworthy AI
---	---

Recommended Readings

Essential Readings:

1. Reshaping Banking with Artificial Intelligence (HKMA white paper).
https://www.hkma.gov.hk/media/eng/doc/key-functions/financial-infrastructure/Whitepaper_on_AI.pdf
2. Beyond the Hype: A Guide to Understanding and Successfully Implementing Artificial Intelligence within Your Business.
<https://www.ibm.com/downloads/cas/8ZDXNKQ4>
3. Motivation for MLOps.
<https://ml-ops.org/content/motivation>
4. What is a Relational Database (RDBMS)?
<https://www.oracle.com/hk/database/what-is-a-relational-database/>
5. Data Warehouse Concepts.
<https://aws.amazon.com/data-warehouse/>
6. OLAP vs. OLTP: What's the Difference?
<https://www.ibm.com/cloud/blog/olap-vs-oltp>
7. What is NoSQL?
<https://www.mongodb.com/nosql-explained>
8. Hadoop Ecosystem and Their Components – A Complete Tutorial.
<https://data-flair.training/blogs/hadoop-ecosystem-components/>
9. Cloud Storage.

- <https://www.ibm.com/cloud/learn/cloud-storage>
10. What is a Data Lake?
https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/?nc1=f_cc
 11. Notes from the AI frontier: Applications and Value of Deep Learning (McKinsey Discussion Paper 2018).
<https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning>
 12. Machine Learning: Bridging Between Business and Data Science.
<https://www.altexsoft.com/whitepapers/machine-learning-bridging-between-business-and-data-science/>
 13. The Simple Guide to Deep Learning.
https://assets.website-files.com/5f1eaa37302cc1d5239e8c91/5f2d82c43857bb21748bf207_simple-guide-to-deep-learning.pdf
 14. Natural Language Processing (NLP).
<https://www.ibm.com/cloud/learn/natural-language-processing>
 15. Some Language Features in Python.
<https://machinelearningmastery.com/some-language-features-in-python/>
 16. NumPy and Pandas Tutorial – Data Analysis with Python.
<https://cloudxlab.com/blog/numpy-pandas-introduction/>
 17. An Introduction to Scikit-Learn: Machine Learning in Python.
<https://www.simplilearn.com/tutorials/python-tutorial/scikit-learn>
 18. Practitioners Guide to MLOps: A Framework for Continuous Delivery and Automation of Machine Learning (Google cloud white paper 2021).
<https://cloud.google.com/resources/mlops-whitepaper>
 19. MLOps Principles.
<https://ml-ops.org/content/mlops-principles>
 20. Machine Learning Operations (MLOps) Framework to Upscale Machine Learning Lifecycle with Azure Machine Learning.
<https://docs.microsoft.com/en-us/azure/architecture/example-scenario/mlops/mlops-technical-paper>
 21. What is DevOps?
https://aws.amazon.com/devops/what-is-devops/?nc1=f_cc

Supplementary Readings

1. Big Data and Machine Learning in Central Banking (BIS working paper).
<https://www.bis.org/publ/work930.htm>
2. What is Data Modeling?

- <https://www.ibm.com/cloud/learn/data-modeling>
3. Data Warehousing: Basics of Relational Vs Star Schema Data Modeling.
<https://medium.com/@daryl.ung/data-warehousing-basics-of-relational-vs-star-schema-data-modeling-75a68eeaf0e3>
 4. NoSQL Databases and Its Types: A Comprehensive Guide 101.
<https://hevodata.com/learn/nosql-databases-and-its-types-a-guide/>
 5. Build a Modern, Unified Analytics Data Platform with Google Cloud (Google Whitepaper 2021).
https://services.google.com/fh/files/misc/googlecloud_unified_analytics_data_platform_paper_2021.pdf
 6. A Refresher on Regression Analysis.
<https://hbr.org/2015/11/a-refresher-on-regression-analysis>
 7. Seven Types of Classification Algorithms.
<https://analyticsindiamag.com/7-types-classification-algorithms/>
 8. Introduction to Classification Algorithms.
<https://dzone.com/articles/introduction-to-classification-algorithms>
 9. Clustering in Machine Learning.
https://training.galaxyproject.org/training-material/topics/statistics/tutorials/clustering_machinelearning/tutorial.html
 10. A Beginner's Guide to Dimensionality Reduction in Machine Learning.
<https://towardsdatascience.com/dimensionality-reduction-for-machine-learning-80a46c2ebb7e>
 11. Five Anomaly Detection Algorithms Every Data Scientist Should Know.
<https://towardsdatascience.com/5-anomaly-detection-algorithms-every-data-scientist-should-know-b36c3605ea16>
 12. Introduction to Recommender Systems.
<https://tryolabs.com/blog/introduction-to-recommender-systems>
 13. Convolutional neural networks: an overview and application in radiology.
<https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9>
 14. Natural Language Processing (NLP): What it is and Why it Matters.
https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html
 15. The Ultimate Guide To Different Word Embedding Techniques In NLP.
<https://www.kdnuggets.com/2021/11/guide-word-embedding-techniques-nlp.html>
 16. Learn Python Programming.
<https://www.programiz.com/python-programming>
 17. Learn Python - Free Interactive Python Tutorial.
<https://www.learnpython.org/>
-

18. PyTorch Tutorial: How to Develop Deep Learning Models with Python.
<https://machinelearningmastery.com/pytorch-tutorial-develop-deep-learning-models/>
19. A Critical Overview of AutoML Solutions.
<https://medium.com/analytics-vidhya/a-critical-overview-of-automl-solutions-cb37ab0eb59e>
20. Choosing the Best MLOps Platform: a Comprehensive Comparison of MLOps Platforms.
<https://valohai.com/mlops-platforms-compared/>
21. Best End-to-End MLOps Platforms: Leading Machine Learning Platforms That Every Data Scientist Need to Know.
<https://neptune.ai/blog/end-to-end-mlops-platforms>

Further Readings

1. 2022 AI Index Report (Stanford University).
<https://aiindex.stanford.edu/report/>