

**深港澳金融科技师（SHMFTPP）**

**一级考试**

**《金融大数据和人工智能》考试大纲**

## 内容说明

本考试大纲适用于深港澳金融科技师 (SHMFTPP) 一级考试，共包含 13 个知识模块，本部分为《金融大数据和人工智能》考试大纲。考试大纲依据各知识点重要程度，将掌握程度要求对应划分为 1、2、3 三个等级。其中 1 级要求理解和基本的记忆，2 级要求记忆并会简单的应用，3 级要求背诵熟练应用掌握。

## 第一部分 金融大数据大纲

### 一、考试目标

1. 大数据概述。包括数据时代和大数据时代的区别，大数据基础概念，大数据分析的特色和挑战，大数据技术概述，大数据的产业应用分布概述，大数据金融应用概述。

2. 大数据技术与产业。熟悉大数据的关键技术框架、存储管理技术、分布式和并行计算技术、以及分析技术。了解大数据产业链、大数据的具体应用，以及大数据与云计算、物联网的关系。

3. 数据的获取和清理。掌握结构化、非结构化和半结构化数据的区别和特点。掌握数据挖掘技术，网络数据挖掘，爬虫，文本数据和图像数据，数据清理等内容。

4. 大数据存储。熟悉传统数据存储和关系数据库。了解分布式文件系统、HDFS、YARN 和 ZooKeeper 等技术，以及 NoSQL 数据库和 HBase 系统的基本原理和应用。

5. 大数据处理和分析。了解 MapReduce、Spark、Storm、图数据和图计算等技术，以及数据可视化和金融数据分布式处理的相关内容。

6. 大数据安全和隐私问题。熟悉数据隐私和数据安全的概念和框架。了解数据分享的原则和规范以及数据交易的前景和挑战。

7. 金融大数据的应用。了解金融科技历史和现状，以 ABCD 为核心的新金融科技、人工智能、区块链和物联网在金融领域的应用等内容。掌握大数据与金融机构转型、金融产品转型等内容。

## 二、考试内容与能力等级

考试内容	能力等级
一、金融大数据分析	
（一）大数据概述	
1. 数据时代和大数据时代的区别	1
2. 大数据基础概念 （1）大数据特则 （2）大数据主要类型	3
3. 大数据分析的特色和挑战：5V 模型	2
4. 大数据技术概述 （1）大数据关键技术框架 （2）大数据存储管理技术 （3）大数据的分布式和并行计算技术 （4）大数据分析技术	2
5. 大数据的产业应用分布概述 （1）大数据产业链 （2）大数据的具体应用 （3）大数据与云计算、物联网的关系	1
6. 大数据金融应用概述	1
（二）数据的获取和清理	
1. 结构化，非结构化数据，半结构化数据的区别和特色	2
2. 数据挖掘技术	2
3. 网络数据挖掘，爬虫	1
4. 文本数据和图像数据	1
5. 数据清理：丢失数据，噪音，错误数据	2
6. 实时数据获取，流数据和批数据，KAFKA	1
7. 实时数据清理	1
8. 金融流数据处理	1
（三）大数据存储	
1. 传统数据存储，关系数据库和关系数据库管理系统	2
2. 关系代数和 SQL 语言基础	1
3. 分布式文件系统 （1）分布式文件系统概念	3

(2) 分布式文件系统的结构 (3) 分布式文件系统的设计需求 (4) 常见的分布式文件系统	
4. HDFS (1) HDFS 的主要概念 (2) HDFS 体系结构 (3) HDFS 的存储原理 (4) HDFS 的数据读写过程	2
5. YARN 和 ZooKeeper	1
6. NoSQL 数据库的几种示例 (1) NoSQL 与关系型数据库的比较 (2) NoSQL 的类型	2
7. HBASE 系统 (1) HBase 数据模型 (2) HBase 运行机制 (3) HBase 的实现原理	2
8. 云存储和云计算	1
9. 金融数据存储实例	1
(四) 大数据处理和分析	
1. MapReduce (1) Hadoop MapReduce 的架构 (2) Hadoop MapReduce 的应用场景	2
2. Spark (1) Apache Spark 生态系统 (2) Apache Spark 架构与原理 (3) Apache Spark 工作流程	1
3. Storm (1) Apache Storm 简介 (2) Apache Storm 的概念 (3) Apache Storm 的框架及其组件	1
4. 图数据和图计算 (1) 图计算概述 (2) 图计算框架概述 (3) 图计算应用场景 (4) 常见图计算模型: Pregel, GraphX, Giraph (5) 图数据存储: Neo4j	1
5. 数据可视化	1
6. 金融数据可视化	1
7. 金融数据分布式处理	1
二、金融大数据应用	
(一) 大数据在金融科技中的应用	
1. 金融科技历史和现状, 以 ABCI 为核心的新金融科技	2
2. 大数据时代的人工智能 (AI) (1) 人工智能的核心: 算法与数据	1

(2) 从软件到硬件，从云端到边缘 (3) 人工智能在金融业的应用	
(二) 数据隐私与数据安全	
1. 数据隐私和数据安全的概念 (1) 数据隐私的定义 (2) 数据安全的定义 (3) 隐私数据类型 (4) 数据安全和隐私保护的重要性	2
2. 数据安全的框架 (1) 数据安全的分类 (2) 数据安全的标准 (3) 数据安全的环节	1
3. 信息加密和数据脱敏（数据屏蔽） (1) 信息加密的概念和意义 (2) 常见的信息加密技术 (3) 数据脱敏概念和方法	1
4. 数据的使用和共享 (1) 数据使用和共享的原则 (2) 数据的归属权和产权 (3) 数据交易	2
5. 数据隐私和数据安全的实践和前沿 (1) 数据隐私安全的监管和法规 (2) 人工智能和大数据时代的数据安全	2
(三) 大数据引导的金融改革	1
1. 大数据与金融机构转型 (1) 大数据与商业银行数字化转型 (2) 大数据与保险机构转型升级 (3) 大数据技术与证券行业 (4) 大数据与信托数字化转型 (5) 大数据与证券投资基金	2
2. 大数据与金融产品转型 (1) 大数据在银行贷款领域的产品创新应用 (2) 大数据对供应链金融的影响和风险 (3) 大数据在票据市场的创新应用 (4) 大数据征信业务的发展 (5) 大数据在保险领域的产品创新应用 (6) 大数据赋能金融交易投资领域	1

## 第二部分 金融人工智能大纲

### 一、考试目标

1. 人工智能概论。了解人工智能的基本定义和四大层次的概念。理解人工智能的主要内容，包括机器学习、知识工程、数据挖掘和知识图谱、自然语言处理、计算机视觉、图神经网络、强化学习、计算机决策和机器人了解人工智能在金融领域的实战应用。

2. 机器学习。熟悉机器学习的表现度量，数据集划分及损失函数的基础概念。掌握机器学习的分类范式。了解常见的机器学习模型。掌握机器学习的过拟合问题、正则化、交叉验证和偏差-方差权衡。了解集成学习，因果学习，迁移学习及可解释性学习的概念。

2. 深度学习。熟悉深度学习的基础概念、深度神经网络架构和向前传播。掌握深度学习进阶的内容，包括正则化、梯度下降的收敛性问题、常见优化方法、批标准化、随机丢弃层、提前终止、参数初始化和超参数搜索空间。掌握基于循环神经网络的时间序列分析和市场预测的实战案例。

3. 自然语言处理。熟悉文本处理基础知识。掌握语言模型和序列到序列模型。了解基于语境的语言模型和注意力机制模型，包括 ELMO、BERT 和 GPT。掌握文本的高维特征发掘和在金融领域实战的应用。

4. 搜索策略与知识推理。了解搜索的含义和状态空间表示掌握盲目搜索方法，包括广度优先搜索、深度优先搜索和一致代价搜索。了解知识表示和知识推理的基本概念，能够描述状态图表示法、问题归约表示法等确定性知识推理方法。

5. AIGC。了解 AIGC 的基本概念、流行背景、主要内容。掌握 AIGC 的核心技术模型，包括深度学习模型和深度生成模型。了解文本生成和图像生成的定义、主要内容、实际应用和相关案例。

6. 强化学习和智能决策。熟悉强化学习的基本概念和算法，包括马可夫决策过程、贝尔曼方程、时序差分算法、Q-learning、策略梯度等，并了解 A2C 和 PPO 算法的初步内容。了解基于强化学习的投资智能体实战，包括 FinRL 和 FinRL-Meta，并能够解释它们的原理和应用。理解博弈决策的基本概念和分类，博弈论和机器博弈决策的基本思路和方法，基于机器学习的博弈决策方法。

7. 金融领域的信息抽取。了解金融领域信息抽取的相关技术和方法。掌握面向金融领域的命名实体识别和关系抽取的方法，以及面向金融领域的时间抽取和事件抽取的方法。

8. 联邦学习和智能问答。理解智能问答系统的概念、分类和构建流程。了解联邦学习的概念、实现流程和在智慧金融、智慧零售等领域的应用。

9. AI 赋能金融。掌握智能问答和机器人流程自动化的概



念和技术框架。了解这两个技术在金融方向上的应用与作用。

## 二、考试内容与能力等级

考试内容	能力等级
一、金融人工智能	
（一）人工智能概述	
1. 基本定义，四大层次的人工智能概念	2
2. 人工智能学派，主要理念和区别	1
3. 人工智能的主要内容：机器学习，知识工程，数据挖掘和知识图谱，自然语言处理，计算机视觉，图神经网络，强化学习，计算机决策，机器人	3
4. 人工智能的挑战	1
5. 人工智能在金融领域实战应用	1
（二）机器学习	
1. 机器学习基础 （1）机器学习的定义 （2）机器学习的表现度量标准：准确度，精度，召回，ROC曲线，AUC，F1值 （3）训练集，测试集，验证集 （3）损失函数的定义和优化	3
2. 机器学习的主要范式 （1）监督学习：分类，回归 （2）无监督学习：聚类 （3）强化学习 （4）半监督学习和自监督学习	2
3. 机器学习的常见方法 （1）逻辑斯蒂回归 （2）朴素贝叶斯 （3）支持向量机 （4）决策树 （5）神经网络	1
4. 机器学习的主要挑战及解决方案 （1）过拟合问题 （2）正则化，交叉验证和偏差-方差权衡	2
5. 常见集成学习方法 （1）引导聚集算法 （2）随机森林算法 （3）提升方法：Adaboost, XGBoost, GBDT	1
6. 因果学习和迁移学习 （1）因果学习相关的概念：因果关系与因果推断 （2）因果监督学习，因果解释	1

(3) 迁移学习相关的概念 (4) 迁移学习的分类和方法	
7. 可解释性学习 (1) 模型可解释性的相关概念 (2) 可解释的模型：回归模型，决策树，决策规则	1
8. 机器学习应用：利用机器学习的因子发掘，用户画像分析	1
(三) 深度学习	
1. 深度学习基础	3
2. 激活函数	1
3. 损失函数	2
4. 深度神经网络架构：向前传播	2
5. 梯度下降和反向传播	1
6. 深度学习进阶 (1) 正则化 (2) 梯度下降的收敛性问题 (3) 常见优化方法：RMSProp, ADAM, momentum (4) 批标准化，随机丢弃层，提前终止 (5) 参数初始化 (6) 超参数和超参数搜索空间 (7) 常见神经网络及其应用：CNN, RNN, LSTM, GRU (8) 深度学习金融实战示例：基于循环神经网络的时间序列分析和市场预测	2
(四) 自然语言处理	
1. 文本处理基础	2
2. 词嵌入算法和词向量：CBOW, Skip-Gram, GloVe	2
3. 语言模型：依存分析，文本 POS 标记	1
4. 序列到序列模型：Bi-LSTM	1
5. 注意力机制模型：Transformer	1
6. 基于语境的语言模型及其应用：ELMO, BERT 系列, Ernie, GPT	1
7. 发掘文本的高维特征，金融领域实战：情感分析，偏见分析，水军/节奏党账户的监测，恶毒语言和网络暴力监控	1
(五) 知识推理，知识图谱与搜索策略	
1. 知识推理与搜索策略概述	2
2. 确定性知识表示方法 (1) 状态图表示法 (2) 问题归约表示法 (3) 谓词逻辑表示法 (4) 语义网络表示法	1
3. 确定性知识表示与推理方法 (1) 图搜索策略 (2) 盲目搜索 (3) 启发式搜索	1

4. 产生式系统及规则演绎系统简介 (1) 产生式系统 (2) 规则演绎系统	1
5. 知识表示下的金融知识图谱应用实例 (1) 知识表示与知识图谱之间的关系 (2) 语义网络知识表示下的金融交易知识图谱应用——反欺诈 (3) 语义网络知识表示下的金融交易用户分析应用——智能投研	1
(六) AIGC	
1. AIGC 概述 (1) 生成式 AI 基本定义 (2) AIGC 流行的原因与背景 (3) AIGC 主要内容与学派：深度学习、预训练模型、文本生成、图像生成、多模态生成 (4) AIGC 的行业应用概述	3
2. 核心技术模型 (1) 深度学习模型：自回归模型，Transformer，BERT，GPT (2) 深度生成模型：VAE，GAN，扩散模型	2
3. 文本生成 (1) 文本生成的定义与主要内容 (2) 聊天机器人 (3) 机器翻译 (4) 多模态文本生成 (5) 文本生成实际应用：ChatGPT	2
4. 图像生成 (1) 图像生成的定义与主要内容 (2) 图像修复 (3) 图像编辑 (4) 图像生成具体应用：Midjourney (5) 视频生成	1
5. AIGC 在金融业的应用与挑战	1
(七) 符号学习与联结学习	
1. 符号学习与联结学习概述 (1) 人工智能的两大流派 (2) 两大流派的发展史 (3) 未来的发展趋势	2
2. 符号学习相关方法与模型 (1) 符号学习起源与专家系统 (2) 逆向演绎与符号回归 (3) 统计学习 (4) 知识图谱	1
3. 联结学习相关方法与模型 (1) 起源与发展阶段	1

(2) 深度学习的兴起 (3) 爆发式发展阶段 (4) 未来的发展	
(八) 强化学习和智能决策	
1. 强化学习概念： 强化学习概念： 智能体， 环境， 奖励函数， 动态过程， 策略和策略函数	2
2. 马可夫过程和马可夫决策过程， 状态价值函数和动作 状态价值函数， 贝尔曼方程	1
3. 时序差分算法和蒙特卡洛算法	2
4. 价值学习, 策略学习, Q-learning	1
5. 策略梯度, on-policy 学习, REINFORCE 算法和 SARSA	1
6. 策略学习进阶： off policy 学习, 经验回放和 actor-critic	1
7. A2C, A3C 和 PPO 算法初步了解	1
8. 基于强化学习的投资智能体实战： FinRL 和 FinRL-Meta	1
9. 博弈论与机器博弈决策	1
10. 机器博弈决策的基本方法	1
11. 博弈决策技术在金融领域的应用 (1) 博弈决策与金融投资	1
(九) 面向金融领域的信息抽取	
1. 面向金融领域的命名实体识别 (1) 面向金融领域的人名识别 (2) 面向金融领域的组织机构名识别	2
2. 面向金融领域的关系抽取	1
3. 面向金融领域的时间抽取 (1) 面向金融领域的时间抽取概述 (2) 面向金融领域的时间抽取方法	1
4. 面向金融领域的事件抽取 (1) 面向金融领域的事件抽取概述 (2) 面向金融领域的事件抽取传统方法 (3) 面向金融领域的事件抽取前沿方法	1
(十) AI 赋能金融	
1. 智能问答 (1) 智能问答概述 (2) 传统与深度技术背景 (3) 交互式问答 (4) 检索式金融智能问答构建流程	1
2. 机器人流程自动化 (Robotic Process Automation) (1) RPA 概述 (2) RPA 技术架构及原理 (3) RPA 与结合 AI 的 应用场景	1
(十一) 联邦学习	
1. 联邦学习概述	1

(1) 背景和挑战 (2) 联邦学习的定义 (3) 学习的分类	
2. 联邦学习与智慧金融 (1) 背景和意义 (2) 实现流程	1
3. 联邦学习与智慧零售 (1) 背景和意义 (2) 实现流程	1