Quiz: Research Agenda

AI-Based Detection of Hedge Fund Fraud

Joerg Osterrieder

Zurich University of Applied Sciences (ZHAW)

2025

Question 1

How many open problems are identified in the research agenda?

a) 5
b) 8
c) 10
d) 12

Question 1

How many open problems are identified in the research agenda?

a) 5
b) 8
c) 10
d) 12

Answer

c) 10

The research agenda identifies exactly 10 open problems (OP1–OP10), organized across three categories: Data challenges, Methodological advances, and Deployment considerations.

Section 6: Research Agenda

Question 2

What are the three categories of open problems?

a) Input/Process/Output
b) Data/Methodological/Deployment
c) Theory/Practice/Policy
d) Design/Build/Test

Question 2

What are the three categories of open problems?

a) Input/Process/Output
b) Data/Methodological/Deployment
c) Theory/Practice/Policy
d) Design/Build/Test

Answer

b) Data/Methodological/Deployment

The 10 open problems are organized into: (1) Data challenges (OP1–OP3), (2) Methodological advances (OP4–OP8), and (3) Deployment considerations (OP9–OP10).

Section 6: Research Agenda

Question 3

Which two problems are identified as critical preconditions for the field?

a) OP2 and OP5
b) OP3 and OP6
c) OP1 and OP4
d) OP8 and OP9

Question 3

Which two problems are identified as critical preconditions for the field?

a) OP2 and OP5
b) OP3 and OP6
c) OP1 and OP4
d) OP8 and OP9

Answer

c) OP1 and OP4

OP1 (Benchmark datasets) and OP4 (Interpretability standards) are critical preconditions. Without accessible data and explainable models, the field cannot progress toward production deployment.

Section 6: Research Agenda

Question 4

What is OP1's two-track approach for creating benchmark datasets?

a) Survey + interview
b) Synthetic data + differential privacy
c) Open source + proprietary
d) Academic + industry

Question 4

What is OP1's two-track approach for creating benchmark datasets?

a) Survey + interview
b) Synthetic data + differential privacy
c) Open source + proprietary
d) Academic + industry

Answer

b) Synthetic data + differential privacy

OP1 proposes generating synthetic fraud datasets while applying differential privacy techniques to protect sensitive information from real cases, enabling research without exposing confidential data.

Section 6.1: Data Challenges

Question 5

What method does OP2 suggest for cross-jurisdictional data integration?

a) Federated learning
b) Blockchain
c) Cloud computing
d) API integration

Question 5

What method does OP2 suggest for cross-jurisdictional data integration?

a) Federated learning
b) Blockchain
c) Cloud computing
d) API integration

Answer

a) Federated learning

OP2 proposes federated learning frameworks that enable model training across multiple jurisdictions without sharing raw data, respecting privacy regulations while leveraging global patterns.

Section 6.1: Data Challenges

Question 6

How many labeled fraud cases typically exist in the literature?

a) 10–20
b) 25–50
c) 50–100
d) 200–500

Question 6

How many labeled fraud cases typically exist in the literature?

a) 10–20
b) 25–50
c) 50–100
d) 200–500

Answer

c) 50–100

The paper identifies that only 50–100 labeled fraud cases exist, creating a severe small-sample problem. OP3 addresses semi-supervised and active learning to work with limited labels.

Section 6.1: Data Challenges

Question 7

What is the name of OP5?

a) Warm-start training
b) Cold-start detection
c) Zero-shot classification
d) Bootstrap analysis

Question 7

What is the name of OP5?

a) Warm-start training
b) Cold-start detection
c) Zero-shot classification
d) Bootstrap analysis

Answer

b) Cold-start detection

OP5 focuses on detecting fraud in funds with limited history (12–24 months of returns), where traditional statistical tests lack power. This requires methods that work with short time series.

Section 6.2: Methodological Advances

Question 8

Which two problems have the lowest feasibility scores?

a) OP1 and OP3
b) OP4 and OP7
c) OP5 and OP6
d) OP2 and OP10

Question 8

Which two problems have the lowest feasibility scores?

a) OP1 and OP3
b) OP4 and OP7
c) OP5 and OP6
d) OP2 and OP10

Answer

d) OP2 and OP10

OP2 (Cross-jurisdictional data) and OP10 (Regulatory alignment) have lowest feasibility due to legal barriers, international coordination challenges, and complex compliance requirements.

Section 6: Priority Matrix