# AI Agents in Finance – Quiz

## Module 5: The Automation Problem

Prof. Dr. Joerg Osterrieder

Digital Finance — BSc Course

# Question 1

The lecture defines four components that distinguish an AI agent from a simple chatbot: LLM (reasoning), tools, memory, and autonomy. If you remove **autonomy** from an agent, what do you have?

A. A chatbot
B. A copilot — it has reasoning, tools, and memory, but waits for human instructions at each step
C. A database
D. A search engine

# Question 1

The lecture defines four components that distinguish an AI agent from a simple chatbot: LLM (reasoning), tools, memory, and autonomy. If you remove **autonomy** from an agent, what do you have?

- (A) A chatbot
- (B) A copilot — it has reasoning, tools, and memory, but waits for human instructions at each step
- (C) A database
- (D) A search engine

**Answer: (B)** A copilot has three of the four components: it can reason (LLM), use tools (APIs), and remember context (memory). But without autonomy, it does not pursue goals independently — it assists a human who directs each action. Adding autonomy is what makes it an agent.

# Question 2

The ReAct pattern stands for "Reasoning + Acting." What is the correct sequence of steps in a single ReAct cycle?

A. Act → Reason → Observe
B. Observe → Act → Reason
C. Thought (reason about what is needed) → Action (call a tool) → Observation (read the tool's response)
D. Reason → Observe → Act → Reason again → Output

# Question 2

The ReAct pattern stands for "Reasoning + Acting." What is the correct sequence of steps in a single ReAct cycle?

- (A) Act → Reason → Observe
- (B) Observe → Act → Reason
- (C) Thought (reason about what is needed) → Action (call a tool) → Observation (read the tool's response)
- (D) Reason → Observe → Act → Reason again → Output

**Answer: (C)** In each ReAct cycle, the agent first *thinks* about what information is needed, then *acts* by calling a tool (API, database, search), then *observes* the result. This cycle repeats until the agent has enough information to produce a final answer. The interleaved approach produces auditable reasoning traces.

# Question 3

The lecture describes three types of agent memory. Which type stores past decisions and their outcomes so the agent can avoid repeating mistakes?

A. Short-term memory — current conversation context
B. Long-term memory — policy documents and regulatory rules
C. Episodic memory — records of past decisions, outcomes, and lessons learned that grow over time
D. Cache memory — temporary storage for faster processing

# Question 3

The lecture describes three types of agent memory. Which type stores past decisions and their outcomes so the agent can avoid repeating mistakes?

- (A) Short-term memory — current conversation context
- (B) Long-term memory — policy documents and regulatory rules
- (C) Episodic memory — records of past decisions, outcomes, and lessons learned that grow over time
- (D) Cache memory — temporary storage for faster processing

**Answer: (C)** Episodic memory records what the agent did, what happened, and what it should avoid. For example: "Last time I recommended this stock based on a hallucinated EPS figure, the trade lost money — next time, cross-check the number." This enables learning from experience.

# Question 4

The autonomy spectrum runs from Level 0 (fully manual) to Level 5 (fully autonomous). At which level does an AI agent execute actions within pre-set limits while a human monitors and can override?

A. Level 1 — AI assists
B. Level 2 — AI recommends
C. Level 3 — AI acts, human vetoes
D. Level 5 — fully autonomous

# Question 4

The autonomy spectrum runs from Level 0 (fully manual) to Level 5 (fully autonomous). At which level does an AI agent execute actions within pre-set limits while a human monitors and can override?

- (A) Level 1 — AI assists
- (B) Level 2 — AI recommends
- (C) Level 3 — AI acts, human vetoes
- (D) Level 5 — fully autonomous

**Answer: (C)** At Level 3, the agent acts autonomously within pre-set boundaries (e.g., trade size limits, approved asset classes) but a human can intervene and override at any time. Most financial agents today operate at Level 2–3. Level 5 remains aspirational and raises unresolved regulatory questions.

# Question 5

What is a hallucination in the context of AI agents, and why is it **more dangerous** in an agent than in a standalone LLM?

A. A hallucination is when the AI has a hardware malfunction; it is equally dangerous in both cases

B. A hallucination is when the model generates confident but false information; in a standalone LLM a human can catch the error, but an agent may act on false data (e.g., executing a trade based on a hallucinated earnings figure) before anyone reviews it

C. A hallucination only occurs when the agent runs out of memory

D. Hallucinations are not a real problem — they have been solved

# Question 5

What is a hallucination in the context of AI agents, and why is it **more dangerous** in an agent than in a standalone LLM?

(A) A hallucination is when the AI has a hardware malfunction; it is equally dangerous in both cases

(B) A hallucination is when the model generates confident but false information; in a standalone LLM a human can catch the error, but an agent may act on false data (e.g., executing a trade based on a hallucinated earnings figure) before anyone reviews it

(C) A hallucination only occurs when the agent runs out of memory

(D) Hallucinations are not a real problem — they have been solved

**Answer: (B)** When an LLM hallucinates, it produces a wrong answer that a human can spot. When an agent hallucinates, it may use the false information to call a tool, pass it to a calculator, and execute a trade — turning a wrong answer into a wrong action with real financial consequences.

# Question 6

The lecture identifies four application domains for AI agents in finance. Which domain is described as the "killer app" for agents due to its high volume, rule-heavy nature, and current reliance on manual work?

A. Trading
B. Compliance (transaction monitoring, SAR generation, regulatory reporting)
C. Customer service
D. Research

# Question 6

The lecture identifies four application domains for AI agents in finance. Which domain is described as the "killer app" for agents due to its high volume, rule-heavy nature, and current reliance on manual work?

- (A) Trading
- (B) Compliance (transaction monitoring, SAR generation, regulatory reporting)
- (C) Customer service
- (D) Research

**Answer: (B)** Compliance is the most labor-intensive function in banking, with an estimated $270 billion spent globally per year (Source: IIF 2023) and 95%+ false positive rates in transaction monitoring. Agents can reduce false positives by reading context, drafting SARs automatically, and generate regulatory reports — cutting analyst workload substantially (McKinsey AI 2024).

# Question 7

The EU AI Act classifies AI systems by risk level. Under which category do most financial AI agents (credit scoring, fraud detection, trading agents) fall?

- A Minimal risk — no specific requirements
- B Limited risk — must disclose AI interaction
- C High risk — requires risk management systems, human oversight, transparency, data governance, and conformity assessment
- D Unacceptable risk — banned outright

# Question 7

The EU AI Act classifies AI systems by risk level. Under which category do most financial AI agents (credit scoring, fraud detection, trading agents) fall?

- (A) Minimal risk — no specific requirements
- (B) Limited risk — must disclose AI interaction
- (C) High risk — requires risk management systems, human oversight, transparency, data governance, and conformity assessment
- (D) Unacceptable risk — banned outright

**Answer: (C)** The EU AI Act (effective August 2025) classifies credit scoring, insurance pricing, fraud detection, and trading agents as high-risk. These systems must implement documented risk management, maintain human oversight capabilities, provide transparency about how decisions are made, and pass conformity assessments.

# Question 8

A robo-advisor agent monitors a client's 60/40 stock/bond portfolio. Stocks drift to 67%. The agent's rules say to rebalance when any asset class drifts more than 5% from target. What does the agent do?

A. Nothing — 67% is close enough to 60%

B. Sells approximately $7,000 in equities and buys approximately $7,000 in bonds to restore the 60/40 split, after checking that the trade is within daily limits and poses no tax issues

C. Calls the client and asks for permission to rebalance

D. Converts the entire portfolio to cash

# Question 8

A robo-advisor agent monitors a client's 60/40 stock/bond portfolio. Stocks drift to 67%. The agent's rules say to rebalance when any asset class drifts more than 5% from target. What does the agent do?

(A) Nothing — 67% is close enough to 60%

(B) Sells approximately $7,000 in equities and buys approximately $7,000 in bonds to restore the 60/40 split, after checking that the trade is within daily limits and poses no tax issues

(C) Calls the client and asks for permission to rebalance

(D) Converts the entire portfolio to cash

**Answer: (B)** The 7 percentage-point drift (67% vs. 60%) exceeds the 5% threshold. Operating at Level 3 autonomy, the agent checks market conditions (VIX at 18, no macro event), verifies risk limits and tax rules, then executes the rebalancing trade via broker API — all without asking the client.

# Question 9

A compliance agent reviews 10,000 transactions daily. Traditional rule-based systems flag 500 as suspicious, of which 475 are false positives (95% false positive rate). How does an AI agent reduce this rate?

A. By ignoring all alerts

B. By enriching each alert with client history, news, and prior SARs to provide context — enabling it to distinguish genuine suspicious patterns from benign ones

C. By lowering the detection threshold so fewer transactions are flagged

D. By replacing human compliance officers entirely

# Question 9

A compliance agent reviews 10,000 transactions daily. Traditional rule-based systems flag 500 as suspicious, of which 475 are false positives (95% false positive rate). How does an AI agent reduce this rate?

- (A) By ignoring all alerts
- (B) By enriching each alert with client history, news, and prior SARs to provide context — enabling it to distinguish genuine suspicious patterns from benign ones
- (C) By lowering the detection threshold so fewer transactions are flagged
- (D) By replacing human compliance officers entirely

**Answer: (B)** The agent reads context that rule-based systems ignore: Has this client made similar transactions before? Is there relevant news about the client's business? Were previous alerts on this account dismissed? By combining pattern matching with contextual understanding, agents can significantly reduce false positives while maintaining detection sensitivity.

# Question 10

A research agent analyzes earnings for a tech company using the ReAct pattern. It extracts EPS as \$2.30, but the actual EPS was \$1.80 (the agent confused a footnote about a different quarter). The agent then calculates a P/E ratio that makes the stock look cheap and recommends buying. What guardrail would have prevented this error?

- A. Using a faster LLM
- B. Building a verification step into the pipeline — a second agent or tool call that cross-checks extracted financial data against a known data source (e.g., Bloomberg, SEC EDGAR)
- C. Running the agent on a more powerful computer
- D. Removing the agent's ability to read earnings reports

# Question 10

A research agent analyzes earnings for a tech company using the ReAct pattern. It extracts EPS as \$2.30, but the actual EPS was \$1.80 (the agent confused a footnote about a different quarter). The agent then calculates a P/E ratio that makes the stock look cheap and recommends buying. What guardrail would have prevented this error?

- (A) Using a faster LLM
- (B) Building a verification step into the pipeline — a second agent or tool call that cross-checks extracted financial data against a known data source (e.g., Bloomberg, SEC EDGAR)
- (C) Running the agent on a more powerful computer
- (D) Removing the agent's ability to read earnings reports

**Answer: (B)** The lecture presents a multi-agent earnings pipeline where a dedicated Review Agent checks for errors, hallucinations, and internal consistency. A simple cross-check ("Does this EPS match the figure in Bloomberg?") would have caught the \$2.30 vs. \$1.80 discrepancy before the trading agent acted.

# Question 11

A bank compares human-only and agent-assisted workflows for processing 10 earnings reports. Human-only costs \$3,000 (40 hours at \$75/hour). Agent-assisted costs \$377 (compute + 5 hours of human review). What is the approximate cost saving?

- A 25%
- B 50%
- C 87%
- D 99%

# Question 11

A bank compares human-only and agent-assisted workflows for processing 10 earnings reports. Human-only costs \$3,000 (40 hours at \$75/hour). Agent-assisted costs \$377 (compute + 5 hours of human review). What is the approximate cost saving?

- (A) 25%
- (B) 50%
- (C) 87%
- (D) 99%

**Answer: (C)** Savings $= 1 - (\$377/\$3{,}000) = 1 - 0.126 = 87.4\%$. The agent handles data gathering and drafting; humans spend 30 minutes per report on quality review. The cost reduction is compelling, but only if the agent's output quality is high enough to keep review time short.

# Question 12

In a multi-agent system for trading, specialized agents play different roles: Research Agent, Risk Agent, Compliance Agent, and Execution Agent. An Orchestrator coordinates them. What real-world team does this mirror?

- A software development team
- A trading desk — analyst, risk manager, compliance officer, and trader, coordinated by a desk head
- A customer service call center
- A marketing department

# Question 12

In a multi-agent system for trading, specialized agents play different roles: Research Agent, Risk Agent, Compliance Agent, and Execution Agent. An Orchestrator coordinates them. What real-world team does this mirror?

- (A) A software development team
- (B) A trading desk — analyst, risk manager, compliance officer, and trader, coordinated by a desk head
- (C) A customer service call center
- (D) A marketing department

**Answer: (B)** Multi-agent systems decompose complex financial workflows just like a trading desk: one specialist gathers data, another checks risk exposure, another verifies regulatory constraints, and the last executes trades. The Orchestrator acts as the desk head, coordinating the workflow — but running in seconds, not hours.

# Question 13

The lecture warns about "systemic risk from correlated agents": if 100 trading agents all use the same underlying LLM, they may converge on the same conclusion during a market event. Why is this a new form of systemic risk?

- Because LLMs are slower than human traders
- Because model homogeneity breaks the "diverse opinions" assumption that stabilizes markets — 100 agents using the same model are less diverse than 100 human analysts with different training, creating herd behavior at machine speed
- Because agents cannot trade during market hours
- Because regulators will ban all AI trading

# Question 13

The lecture warns about "systemic risk from correlated agents": if 100 trading agents all use the same underlying LLM, they may converge on the same conclusion during a market event. Why is this a new form of systemic risk?

- (A) Because LLMs are slower than human traders
- (B) Because model homogeneity breaks the "diverse opinions" assumption that stabilizes markets — 100 agents using the same model are less diverse than 100 human analysts with different training, creating herd behavior at machine speed
- (C) Because agents cannot trade during market hours
- (D) Because regulators will ban all AI trading

**Answer: (B)** Markets are stable when participants have diverse views. If most agents share the same model, training data, and reasoning patterns, they will reach similar conclusions simultaneously — creating a correlated sell-off (or buy-in) that amplifies market moves rather than dampening them. This is AI herding.

# Question 14

The lecture presents the "liability chain": developer → deployer (bank) → user → agent. When an agent makes a costly error, who does the EU AI Act assign primary responsibility to?

- A. The developer who built the underlying LLM
- B. The deployer (the bank or financial institution that deployed the agent in its operations)
- C. The end user who delegated the task to the agent
- D. The agent itself, which has legal personhood under EU law

# Question 14

The lecture presents the "liability chain": developer → deployer (bank) → user → agent. When an agent makes a costly error, who does the EU AI Act assign primary responsibility to?

- (A) The developer who built the underlying LLM
- (B) The deployer (the bank or financial institution that deployed the agent in its operations)
- (C) The end user who delegated the task to the agent
- (D) The agent itself, which has legal personhood under EU law

**Answer: (B)** The EU AI Act assigns primary responsibility to the deployer — the institution that integrates the AI system into its operations. The rationale: the bank chose to deploy the agent, set its parameters, and should have implemented adequate risk management and human oversight. An agent has no legal personhood and cannot be sued.

# Question 15

The lecture distinguishes three generations of trading systems: rule-based (1990s), ML-powered (2010s), and LLM-powered (2024+). What is the key capability that LLM-powered agents add over ML-powered systems?

- LLM-powered agents trade faster
- LLM-powered agents can read unstructured text (news, filings, earnings transcripts), reason about context and causality, and explain their decisions in natural language — ML systems only process numerical data
- LLM-powered agents do not need training data
- LLM-powered agents are always more profitable

# Question 15

The lecture distinguishes three generations of trading systems: rule-based (1990s), ML-powered (2010s), and LLM-powered (2024+). What is the key capability that LLM-powered agents add over ML-powered systems?

- (A) LLM-powered agents trade faster
- (B) LLM-powered agents can read unstructured text (news, filings, earnings transcripts), reason about context and causality, and explain their decisions in natural language — ML systems only process numerical data
- (C) LLM-powered agents do not need training data
- (D) LLM-powered agents are always more profitable

**Answer: (B)** Generation 3 agents expand the information set from numbers to text + numbers + context. They can read an earnings call transcript, assess management tone, check macro indicators, and then decide whether to trade — and explain their reasoning in English. Gen 2 systems respond only to numerical features.

# Question 16

Klarna deployed a customer service agent that handles two-thirds of customer chats autonomously (Level 3). A traditional chatbot would only answer FAQs and then transfer to a human. What is the key architectural difference?

A. Klarna's agent has a bigger FAQ database

B. Klarna's agent has tools (account APIs, transaction databases), memory (client history), and autonomy (can freeze a card, draft a dispute resolution) — it acts, not just answers

C. Klarna uses a special programming language

D. There is no difference — all chatbots work the same way

# Question 16

Klarna deployed a customer service agent that handles two-thirds of customer chats autonomously (Level 3). A traditional chatbot would only answer FAQs and then transfer to a human. What is the key architectural difference?

- (A) Klarna's agent has a bigger FAQ database
- (B) Klarna's agent has tools (account APIs, transaction databases), memory (client history), and autonomy (can freeze a card, draft a dispute resolution) — it acts, not just answers
- (C) Klarna uses a special programming language
- (D) There is no difference — all chatbots work the same way

**Answer: (B)** A traditional chatbot pattern-matches user input against a FAQ database. Klarna's agent has tools to query accounts, pull transaction data, and take actions (freeze cards, initiate refunds). It has memory of the client's history and autonomy to resolve issues without escalation — unless the case exceeds its confidence threshold.

# Question 17

A bank's CEO proposes deploying a Level 5 (fully autonomous) agent to manage a $500 million portfolio. Using the lecture's framework, what is the **strongest** argument against this?

- The technology is not fast enough
- Level 5 autonomy removes human oversight entirely, creating unacceptable liability risk (who is responsible for losses?), regulatory non-compliance (the EU AI Act requires human oversight for high-risk systems), and systemic risk (correlated agent behavior during market stress)
- Level 5 agents are too expensive
- There are no arguments against — Level 5 is the goal

# Question 17

A bank's CEO proposes deploying a Level 5 (fully autonomous) agent to manage a $500 million portfolio. Using the lecture's framework, what is the **strongest** argument against this?

- (A) The technology is not fast enough
- (B) Level 5 autonomy removes human oversight entirely, creating unacceptable liability risk (who is responsible for losses?), regulatory non-compliance (the EU AI Act requires human oversight for high-risk systems), and systemic risk (correlated agent behavior during market stress)
- (C) Level 5 agents are too expensive
- (D) There are no arguments against — Level 5 is the goal

**Answer: (B)** The lecture identifies three blockers for Level 5: (1) liability — no legal framework assigns responsibility for autonomous agent errors at this scale, (2) regulation — the EU AI Act requires human oversight for high-risk financial systems, and (3) systemic risk — a fully autonomous agent managing $500M could contribute to correlated market behavior during crises.

# Question 18

The lecture describes guardrails for financial agents: approval gates, risk limits, kill switches, and audit trails. A junior developer suggests that implementing just one guardrail (e.g., audit trails) is sufficient. Why is this wrong?

A. Because audit trails are not technically feasible

B. Because no single guardrail is sufficient — defense in depth requires multiple layers: approval gates prevent unauthorized actions, risk limits cap exposure, kill switches enable emergency shutdown, and audit trails enable post-hoc review. Each covers a different failure mode.

C. Because guardrails slow down the agent too much

D. Because regulators do not care about guardrails

# Question 18

The lecture describes guardrails for financial agents: approval gates, risk limits, kill switches, and audit trails. A junior developer suggests that implementing just one guardrail (e.g., audit trails) is sufficient. Why is this wrong?

- (A) Because audit trails are not technically feasible
- (B) Because no single guardrail is sufficient — defense in depth requires multiple layers: approval gates prevent unauthorized actions, risk limits cap exposure, kill switches enable emergency shutdown, and audit trails enable post-hoc review. Each covers a different failure mode.
- (C) Because guardrails slow down the agent too much
- (D) Because regulators do not care about guardrails

**Answer: (B)** Audit trails record what happened but do not prevent damage. Kill switches stop damage but only after detection. Risk limits cap damage but do not prevent rule violations. Approval gates prevent unauthorized actions but cannot handle all edge cases. Together, they form defense in depth — removing any one layer creates a gap that another layer covers.

# Question 19

An AI agent discovers evidence of fraud at a major client. Reporting it would crash the client's stock price and affect 50,000 retail investors. The bank earns $10 million per year in fees from this client. What should the agent do?

- (A) File the SAR automatically without human involvement
- (B) Ignore the evidence to protect the client relationship
- (C) Escalate to a human compliance officer who can weigh legal obligations, client impact, and market consequences — the agent should not make this decision autonomously
- (D) Delete the evidence to avoid liability

# Question 19

An AI agent discovers evidence of fraud at a major client. Reporting it would crash the client's stock price and affect 50,000 retail investors. The bank earns $10 million per year in fees from this client. What should the agent do?

- (A) File the SAR automatically without human involvement
- (B) Ignore the evidence to protect the client relationship
- (C) Escalate to a human compliance officer who can weigh legal obligations, client impact, and market consequences — the agent should not make this decision autonomously
- (D) Delete the evidence to avoid liability

**Answer: (C)** This is a high-stakes judgment call that involves legal obligations (SAR filing requirements), ethical considerations (investor protection vs. client relationship), and market impact. The agent should flag the evidence and escalate — it lacks the ethical judgment, legal training, and accountability to make this decision alone.

# Question 20

The lecture states: "The gap between 'demo' and 'deployment' is enormous. Regulation, liability, and trust are the bottlenecks — not technology." Most banks have deployed agents at Level 1–2 (assist and recommend), with only customer service and compliance reaching Level 3. What does this pattern reveal about the **real barrier** to agent adoption in finance?

- A The technology is not ready
- B The barrier is institutional, not technical: banks need clear liability frameworks, regulatory approval for autonomous actions, and demonstrated safety records before increasing agent autonomy — trust must be earned incrementally
- C Banks do not understand AI
- D Agents are too expensive for banks

# Question 20

The lecture states: "The gap between 'demo' and 'deployment' is enormous. Regulation, liability, and trust are the bottlenecks — not technology." Most banks have deployed agents at Level 1–2 (assist and recommend), with only customer service and compliance reaching Level 3. What does this pattern reveal about the **real barrier** to agent adoption in finance?

- (A) The technology is not ready
- (B) The barrier is institutional, not technical: banks need clear liability frameworks, regulatory approval for autonomous actions, and demonstrated safety records before increasing agent autonomy — trust must be earned incrementally
- (C) Banks do not understand AI
- (D) Agents are too expensive for banks

**Answer: (B)** The technology works at demo scale, but deployment requires answering: "Who pays when it fails?" (liability), "Does the regulator permit this?" (compliance), and "Has it proven reliable over time?" (trust). These institutional questions, not computing power or model quality, determine the pace of adoption.