AI Agents in Finance

The Robot That Trades While You Sleep

Day 7 of 10

Prof. Jörg Osterrieder

BSc Seminar: Digital Finance

Spring 2026

When AI Resolves a $100M Bet

Polymarket uses LLM-based oracles (GPT-4 + UMA protocol) to resolve prediction markets.

- **89% accuracy** across 1,660 resolved markets
- But: an LLM once **misread a headline** about a merger and incorrectly resolved a $500K market

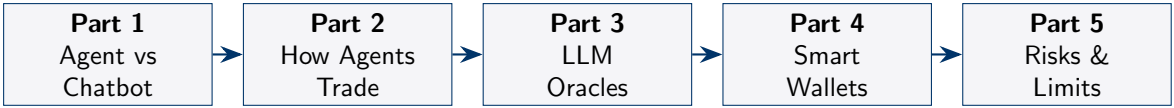
89%
accuracy
across 1,660 markets

1 error =
$500K resolved
incorrectly

The Question

Should we trust AI with financial decisions?
If yes — under what conditions?
If no — what is the alternative?

Today's Agenda



Outline

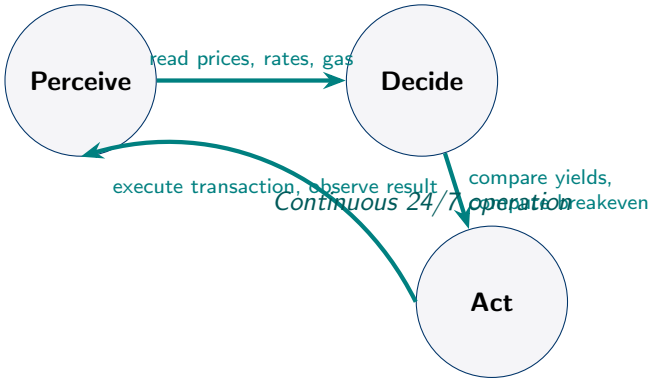
Chatbot vs. Agent: The Key Difference

	Chatbot (ChatGPT)	AI Agent
What it does	Answers questions	**Takes actions**
Who initiates	You ask, it replies	It acts on its own
Uses tools?	No (text only)	Yes (APIs, blockchain)
Runs 24/7?	Only when you ask	**Continuously**
Example output	"BTC is at $60K"	*Moves $50K to higher-yield pool*

One-Liner

A chatbot **talks** to you. An agent **does things for** you.
If ChatGPT is a translator, a DeFi agent is an **autonomous portfolio manager**.

The Agent Loop: Perceive – Decide – Act



Continuous 24/7 operation

The agent repeats this loop **every few minutes**, adjusting its strategy based on new data — while you sleep.

Example: Yield Optimizer Agent

The Setup

Agent monitors 15 DeFi lending protocols across 3 blockchains.
Your \$50,000 USDC is currently in Aave on Ethereum at 2.0% APY.

Agent's calculation (every 5 minutes):

1. **Perceive:** Best rate found: Morpho on Base at 5.1% APY
2. **Decide:**
   - Improvement: $5.1\% - 2.0\% = 3.1\%$ on \$50K = \$1,550/year extra
   - Switch cost: \$3.52 (gas + bridge fee)
   - Breakeven: $\$3.52/(\$1{,}550/365) = 0.8$ days
3. **Act:** Withdraw from Aave $\rightarrow$ Bridge to Base $\rightarrow$ Deposit in Morpho

Result: Extra \$1,547/year. Executed at 3:14 AM automatically.

Outline

How Agents Think Step by Step: The ReAct Pattern

Thought: ETH dropped 8%. Let me check my Aave health factor.

Action: Call aave.getHealthFactor() → Result: **1.15** (below 1.3 threshold!)

Thought: Health factor too low. I need to repay $2,000 debt to restore safety.

Action: Call aave.repay(USDC, 2000) → Result: New health factor: **1.32**

Thought: Position secured. Setting price alert for further drops.

Key: The AI *thinks step by step* and *checks its work* by querying real blockchain data between each thought.

Why Agents Beat Humans in DeFi

Task	Human	Agent
Check 15 lending rates	30 min	2 seconds
Compare gas across chains	10 min	0.5 seconds
React to price crash	Minutes to hours	Milliseconds
Available	$\sim$8 hrs/day	**24/7/365**
Emotional decisions	Frequent	Never
Transaction errors	Possible	Rare (but different risks)

DeFi runs 24/7. If you sleep 8 hours, you miss 33% of the market. An agent never sleeps.

Outline

LLM Oracles: AI Reads the News to Settle Bets

The Problem

Traditional oracles (like Chainlink) report **numbers**: prices, temperatures.
But some questions need **judgment**: "Did the EU approve MiCA amendments?"

The Solution: LLM Oracle

An AI model **reads official sources** (EU press releases, Reuters) and makes a judgment call.
Used by Polymarket + UMA protocol.

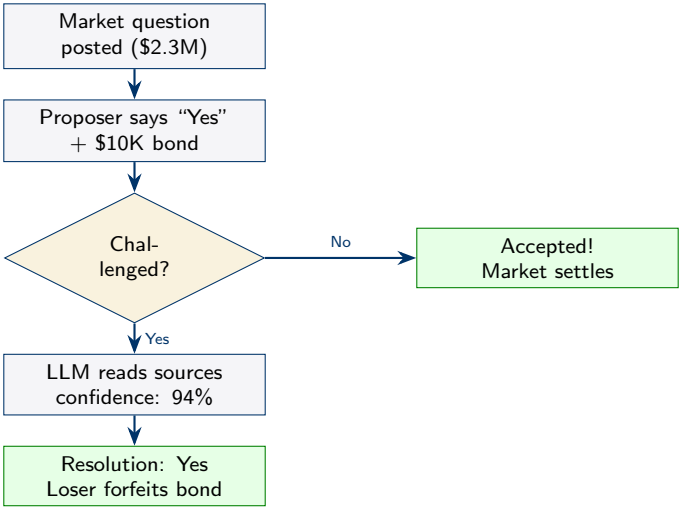
Chainlink can answer:

- What is the ETH price? ($3,000)
- What is the BTC/USD rate?

LLM oracle can answer:

- Did the EU approve MiCA?
- Was the election certified?
- Did Hurricane X hit Florida?

How LLM Resolution Works



89% Accurate — Is That Good Enough?

The good:

- 89% accuracy across 1,660 markets
- Faster than human judges (minutes vs. days)
- Cheaper than hiring arbitration panels
- Scales to thousands of markets

The bad:

- 11% error rate on \$100M+ = millions lost
- AI misread a merger headline once
- Stale training data can miss recent events
- Adversaries can plant fake news to fool the AI

Thought Experiment

Would you trust an AI judge that is right 89% of the time with your \$10,000 bet?

What about 95%?
What about 99%?

Where is your threshold?

Outline

Today's Wallets Are "Dumb Keys"

Current wallet (EOA):

- Sign one transaction at a time
- Must pay gas in ETH only
- Lose your seed phrase = lose everything
- Cannot delegate anything to an AI
- 3 actions = 3 transactions = 3 fees

Smart wallet (EIP-7702):

- **Batch** multiple actions in one click
- Pay gas in **any token** (USDC, etc.)
- **Social recovery** (friends help you recover)
- **Delegate** to AI with spending limits
- 3 actions = 1 transaction = 1 fee

Analogy

Upgrading from a **physical key** (one lock, one door) to a **smart lock** (access rules, time limits, guest codes).

Session Keys: Safe Delegation to AI Agents

How It Works

You create a "session key" for your AI yield agent:

Permission	Setting
Allowed actions	Deposit/withdraw on Aave, Compound, Morpho
Spending limit	$1,000 per day
Expiration	30 days
NOT allowed	Transfer to external addresses

If agent is hacked: max loss = $1,000 (daily limit). Revoke the key.
Your main private key: never shared with the agent.
Result: AI trades for you 24/7 within guardrails *you* define.

Before & After: User Experience

Before Account Abstraction

1. Approve USDC (sign + gas in ETH)
2. Swap USDC for ETH (sign + gas)
3. Deposit ETH in Aave (sign + gas)

Total: 3 transactions, $15 in gas, 3 manual signatures.

No ETH for gas? **You are stuck.**

After Account Abstraction

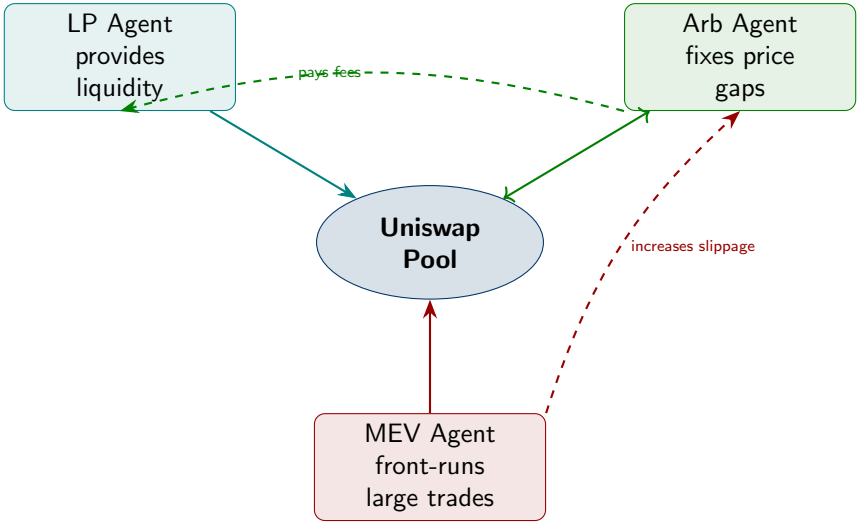
1. All 3 actions batched into ONE

Total: 1 signature, $5 gas (paid in USDC), someone else can even sponsor the gas.

No ETH? **No problem.**

Outline

When Multiple AIs Trade the Same Market



Cooperative: Arb Agent pays fees to LP Agent

Outline

Three Ways AI Agents Can Fail

Hallucination

AI generates **plausible but false** output.

Example: Agent asked for Uniswap's address on Arbitrum. LLM returns the **Ethereum** address. Sends $50K to wrong chain.

Alignment Failure

Agent optimizes the **wrong goal**.

Example: "Maximize yield" leads agent to deposit $100K in unaudited "SuperYield" protocol offering 500% APY. Rug pull.

Adversarial Attack

Someone **tricks** the agent.

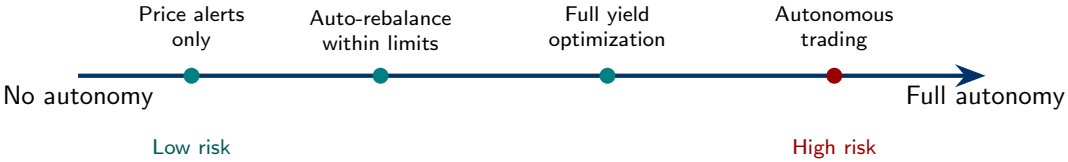
Example: Attacker pumps a worthless token on a DEX. Agent reads fake price as real. Buys $50K of worthless tokens.

How to Make Agents Safer

Risk	Mitigation	Implementation
Hallucination	Verify against multiple sources	Cross-check 3+ data feeds
Alignment	Add risk constraints to goals	"Max yield *from audited protocols only*"
Adversarial	Use multiple oracles	Require 3/5 oracle agreement
All of the above	Session key limits	\$1K/day spending cap
All of the above	Human-in-the-loop	Require approval >\$10K

Key principle: More autonomy = more risk. Start with tight guardrails, loosen as trust builds.

The Trust Spectrum



Question for the class: Where on this spectrum would you trust an AI with your money?

Outline

Exercise: Design a Simple Yield-Checking Agent

Your Task

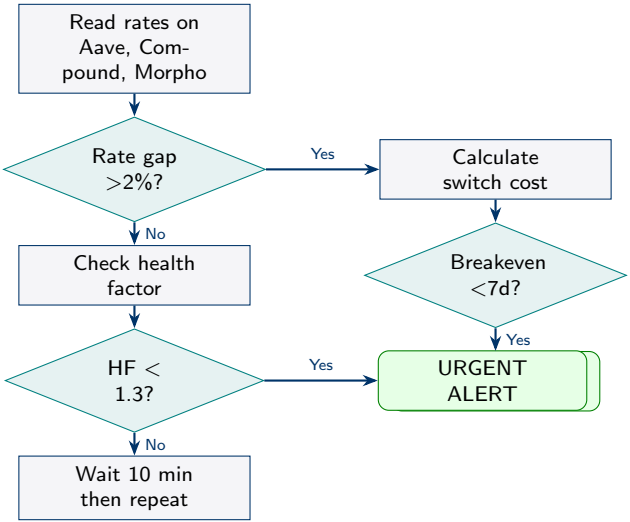
Design a flowchart for an agent that monitors DeFi yields and alerts you when to switch. **No coding required** — just draw the logic.

Your agent should:

1. Check lending rates on 3 protocols every 10 minutes
2. If a rate is $>2\%$ higher than your current rate:
   - Calculate switching cost (gas + bridge fees)
   - Calculate breakeven time
   - If breakeven < 7 days: recommend switching
3. If your health factor drops below 1.3: send urgent alert

Deliverable: A flowchart with decision diamonds for each "if" statement. Use: Perceive $\rightarrow$ Decide $\rightarrow$ Act structure.

Example Flowchart Structure



Day 7: Key Takeaways

1. **AI agents** go beyond chatbots: they perceive, decide, and act autonomously in DeFi markets
2. Agents follow the **Perceive–Decide–Act** loop, checking real blockchain data at every step
3. **LLM oracles** can resolve subjective questions ("Did X happen?") but 89% accuracy is not 100%
4. **Smart wallets** (EIP-7702) enable safe delegation: session keys with spending limits and expiration
5. **Risks are real:** hallucination, alignment failure, adversarial attacks. Guardrails are essential.

Discussion Questions

1. Would you trust an AI agent with $10K? $100K? $1M? Where is your comfort threshold?
2. What happens when thousands of AI agents compete in the same market? More efficient or more fragile?
3. Should AI agents in finance be regulated? Who is responsible when an agent loses money?
4. How is an AI trading agent different from algorithmic trading that banks have done for decades?

Next Time: Privacy in Crypto

Day 8: Proving Facts Without Revealing Secrets

- Zero-knowledge proofs: the colorblind balls story
- SNARKs vs. STARKs: two flavors of ZK
- ZK-rollups: making Ethereum 100$\times$ cheaper
- Private DeFi: trading without everyone watching
- The privacy vs. compliance debate

References I