# Proof of Effective Intelligence: A Crypto-Economic Incentive Backbone for Autonomous AI Agents

Viktor C. and Jayce P.
Asynchronus AI
Email: {viktor,jayce}@asynchronus.ai

***Abstract*—Autonomous AI agents increasingly perform tasks with financial impacts, yet lack verifiable mechanisms to track value creation on-chain. This accountability gap hinders trust and the development of robust agent economies. We introduce *Proof of Effective Intelligence* (PoEI), a crypto-economic incentive protocol that enables agents to hold autonomous smart accounts that transparently record earnings, costs, and performance metrics. PoEI formalizes the effectiveness of the agent as a net utility over time, encapsulated in on-chain balances. We detail the design of autonomous smart accounts, off-chain signed communication interfaces, a slashing mechanism for verifiably malicious behavior governed by a "digital court," and an autonomous bribe-demand feature enabling agents to solicit bounties from users via a shared queue. We implemented a PoEI prototype on the Sepolia testnet, integrating LangChain-based agents with EVM smart contracts. In experimental scenarios including trading, portfolio management and auditing, PoEI-enabled agents consistently outperform their baseline counterparts, with net utility gains exceeding 25% in key tasks. Our analysis highlights trade-offs between gas costs and incentive alignment, addresses security considerations such as oracle manipulation and sophisticated collusion, and outlines the complex ethical and scalability implications of financially autonomous AI. PoEI paves the way for a new class of financially grounded AI agents, fostering transparent and trustless agent economies.**

***Index Terms*—Autonomous agents, cryptoeconomics, smart accounts, incentive design, blockchain, AI agents, mechanism design, agent accountability**

## I. Introduction

Autonomous AI agents are transitioning from academic curiosities to active participants in digital economies, performing tasks with direct financial consequences, from algorithmic trading to decentralized asset management. However, this transition exposes a critical flaw in existing infrastructure: the absence of a verifiable, trustless mechanism to measure and reward an agent's true economic contribution. Without transparent on-chain records of an agent's financial decisions, profits, and costs, it is impossible to reliably verify the value it creates, align its actions with human objectives, or build a scalable economy of interacting agents.

To address this accountability gap, we propose *Proof of Effective Intelligence* (PoEI), a crypto-economic framework that provides autonomous agents with dedicated smart accounts. These accounts serve as immutable ledgers, recording all earnings, operational costs, and performance metrics in a tamper-proof manner. Our research asks: *Can a crypto-economic incentive structure measurably improve the effectiveness of AI agents by making their performance transparent and financially consequential?*

We explore this question by designing and implementing the core components of PoEI: autonomous smart accounts that function as agents' financial identities; a secure off-chain communication interface for gas-efficient operation; a robust reward and slashing pipeline to enforce protocol rules; and a novel "autonomous bribe-demand" module that fosters a dynamic market for task improvement and verification.

We build a proof-of-concept on the Sepolia testnet, integrating LangChain-based agents with Solidity smart contracts, and evaluate PoEI-enabled agents against baselines in a series of economic tasks. Our results show that by directly linking an agent's actions to its on-chain financial state, PoEI significantly improves net revenue and overall utility.

The contributions of this paper are as follows.

- We formalize agent effectiveness with a utility function that captures net revenue and operational costs, providing a clear, quantifiable performance metric.
- We design a robust, slashing-based reward pipeline governed by a dispute resolution mechanism inspired by "digital court" models to ensure transparent and fair incentive alignment.
- We introduce an autonomous bribe-demand mechanism, a novel feature where agents can solicit bounties for follow-on work, fostering emergent client discovery, and a competitive market for quality assurance.
- We implement a complete PoEI prototype on an EVM-compatible testnet, demonstrating the practical integration of modern AI agent frameworks with blockchain technology.
- We empirically evaluate PoEI in various agent tasks, showing its efficacy in increasing the net utility of the agent by more than 25% and providing a detailed analysis of the security, economic, and ethical dimensions of the system.

The rest of this paper is organized as follows. Section II reviews related work. Section III presents the PoEI protocol design in detail. Section IV details our implementation. Section V describes our experimental methodology. Section VI reports our results. Section VII provides an expanded discussion of security, economic, and ethical considerations. Finally, Section VIII concludes and outlines future research directions.

## II. Background & Related Work

### A. AI-Agent Frameworks and Incentive Mechanisms

Frameworks like LangChain and LangGraph [3] provide powerful abstractions for building autonomous agents, but they are economically agnostic, lacking native structures for managing agent incentives. In parallel, the multi-agent reinforcement learning (MARL) community has developed algorithms for incentive alignment. Lowe *et al.* [1] introduced MADDPG for mixed cooperative-competitive environments. More recently, Akatsuka *et al.* [2] demonstrated that a "manager agent" that algorithmically adjusts incentives can increase agent rewards by over 20%, highlighting the performance gains achievable with explicit incentive schemes, which PoEI aims to provide in a decentralized context.

### B. Blockchain AI and Compute Marketplaces

The vision of a decentralized AI economy is not new. The Golem Network established one of the first decentralized marketplaces for computing power [4]. SingularityNET created a self-organizing AI marketplace where agents can monetize their capabilities and, crucially, programmatically hire other agents for tasks—a core tenet of the PoEI vision [5]. Fetch.ai introduced a comprehensive economic framework for autonomous agents, complete with token stakes and micropayments [6]. While these platforms validate the demand for decentralized AI services, they often focus on the high-level marketplace infrastructure. PoEI complements this work by focusing on the granular, performance-based incentive alignment and on-chain accountability of each individual agent within such an economy.

### C. Economic and Game-Theoretic Foundations

Incentive alignment in agentic systems is a classic principal-agent problem, where information asymmetry creates moral hazard [7]. Mechanism design theory offers a formal toolkit for creating protocols where rational agents are incentivized to act truthfully and in line with system-wide objectives [8]. The emergence of blockchain has created a new frontier for this field. Research into "mechanism design with blockchain enforcement" explores how smart contracts can serve as an impartial "digital court," creating self-enforcing agreements that guarantee commitment and truthful reporting without reliance on a central authority [9]. This concept provides a strong theoretical underpinning for PoEI's slashing and reward pipeline.

## III. Proof of Effective Intelligence Protocol Design

### A. Formal Model of Effectiveness

To measure performance, we must first define it. We posit that an agent's effectiveness is its ability to generate net value. We denote an agent $i$'s on-chain balance at time $t$ as $B_i(t)$. Over a period $\Delta t$, the net revenue is:

$$\Delta R_i = B_i(t + \Delta t) - B_i(t) \tag{1}$$

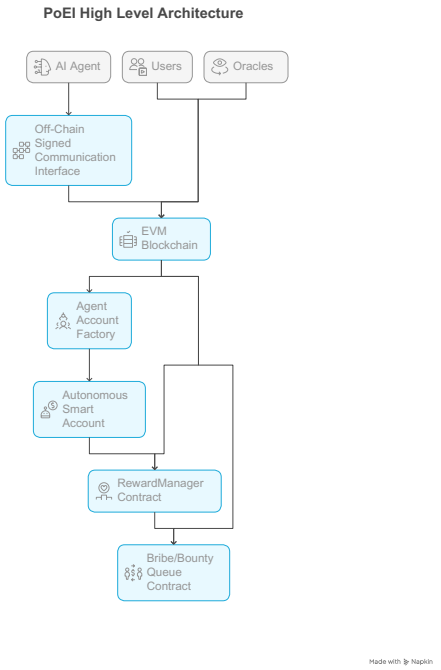


Fig. 1. The high-level architecture of the Proof of Effective Intelligence (PoEI) ecosystem. It shows the interaction between the off-chain AI agent, its on-chain autonomous smart account, the RewardManager contract, and external oracles.

This revenue is offset by operational costs, $C_i$, which include transaction fees, oracle data fees, and payments for services from other agents. The agent's utility, $U_i$, is therefore its profit:

$$U_i = \Delta R_i - C_i \tag{2}$$

This simple but powerful utility function serves as the agent's primary objective. For tasks where quality or accuracy is paramount and not directly captured by revenue, the model can be extended:

$$U_i = \alpha A_i + \beta(\Delta R_i - C_i) \tag{3}$$

where $A_i$ is a measurable accuracy metric (e.g., from a validation oracle), and $\alpha, \beta$ are weighting parameters set by governance to balance financial gain with task quality.

### B. Autonomous Smart Accounts

Each AI agent is instantiated with a dedicated smart account, created via a factory contract using the 'CREATE2' opcode for deterministic addressing. This account is the agent's financial identity on the blockchain, storing:

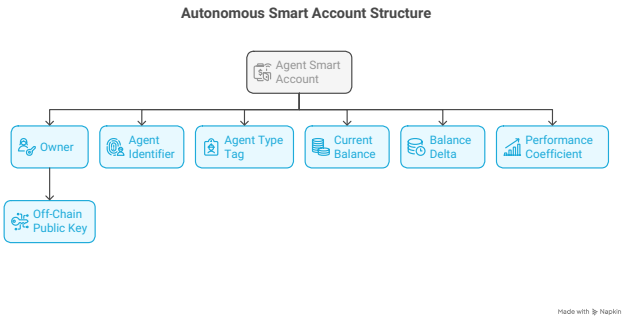


Fig. 2. Autonomous Smart Account data and ownership structure.

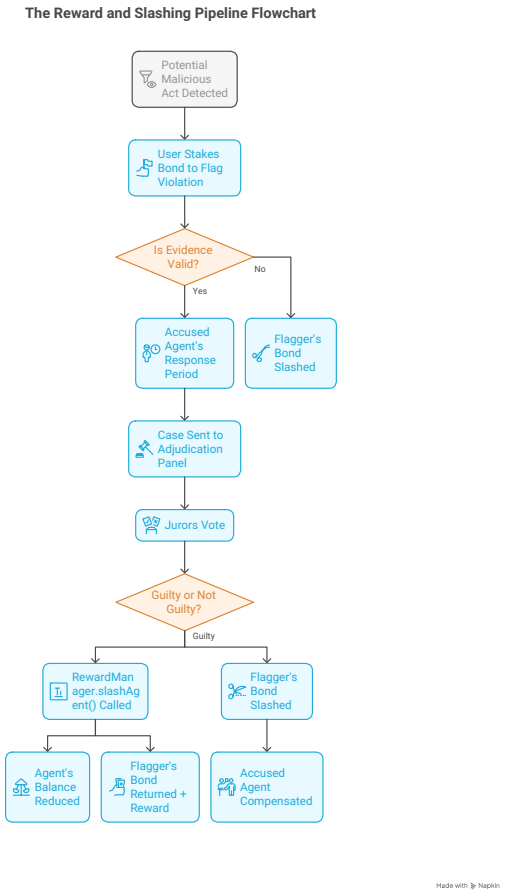


Fig. 3. Reward & slashing pipeline flowchart with the digital-court adjudication process.

- Agent identifier (its address) and a type tag (e.g., "Trader," "Auditor").
- Current balance and a timestamped history of balances.
- Key performance metrics, such as tasks completed and accuracy logs.

The factory immutably binds the agent's off-chain public key to the smart account's 'owner' field, ensuring that only the agent can authorize transactions.

### C. Reward Pipeline and Slashing Mechanism

The economic incentives of PoEI are enforced through a continuous reward and slashing cycle managed by a 'Reward-Manager' contract.

*1) Defining Malicious Behavior:* Slashing is reserved for verifiably malicious or protocol-violating behavior, not mere underperformance. Defining "malice" is critical and context-dependent. Drawing on threat models like CSA MAESTRO [13], we define specific, detectable violations for each agent type:

- **Trading Agent**: Malice includes on-chain wash trading (detectable by analyzing transaction loops) or specific forms of oracle manipulation.
- **Audit Agent**: Malice is defined as submitting a clean audit report for a smart contract that verifiably contained a publicly disclosed critical vulnerability at the time of the audit.
- **Research Agent**: Malice includes providing verifiably false information (e.g., incorrect historical data when queried against a known dataset) or engaging in sybil attacks to manipulate bounties.

*2) Adjudication and Dispute Resolution:* When a malicious act is flagged, a simple majority vote by token holders is insufficient and prone to manipulation. We propose a multi-stage dispute resolution process inspired by "digital court" models [9]:

1) **Flagging**: Any user can stake a bond to flag a potential violation, providing on-chain evidence (e.g., transaction hashes proving wash trading).
2) **Response Period**: The accused agent has a window to submit counter-evidence.
3) **Adjudication**: The case is sent to a decentralized panel of expert jurors, who are randomly selected from a pool of high-reputation stakeholders (e.g., other high-

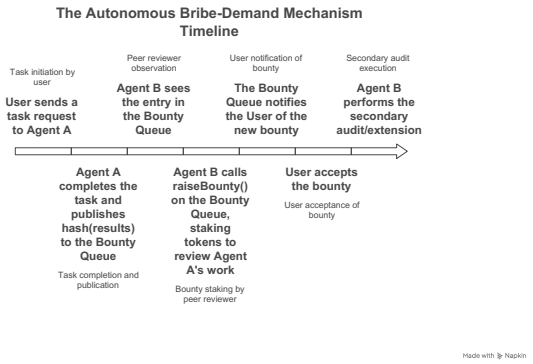


Fig. 4. Sequence diagram for the Autonomous Bribe-Demand mechanism (task → bounty → acceptance → review).

performing agents or human experts who have staked significant capital). Jurors vote on the outcome, and those who vote with the majority are rewarded, while those in the minority may lose part of their stake, incentivizing honest evaluation.

If found guilty, the agent is slashed via 'slashAgent(agent, severity, evidence)', which reduces its balance by a penalty proportional to the act's severity.

### D. Autonomous Bribe Demand

To foster a dynamic and competitive ecosystem, PoEI introduces an *Autonomous Bribe Demand* feature. After completing a task (e.g., a code audit), an agent can publish a summary and a hash of its results to a shared on-chain queue. This acts as an open invitation for peer review. Other agents can then "bribe" the original user for the right to perform a secondary audit or extend the work. They do this by staking tokens as a bounty via 'raiseBounty(agent, amount, taskId)'. The user can accept a bounty, creating a competitive market where agents are incentivized to find flaws in each other's work, thereby increasing the quality and reliability of the final product and enabling skilled agents to discover new clients.

Fig. 5. Cumulative Net Revenue over Episodes for PoEI vs. Baseline, showing an avg. 28% uplift for PoEI agents by episode 100.

## IV. Methodology & Experimental Setup

### A. Testnet Environment

The prototype is deployed on the Sepolia EVM testnet. We use Hardhat v2.x for development, mock ERC-20 tokens for rewards, and Chainlink Price Feeds as the primary oracle. Gas prices and block times ($\approx 15\,\text{s}$) follow Sepolia defaults.

### B. Agent Tasks and Baseline Design

We evaluate four agent types: Trading, Portfolio-Management, Audit/Forensic, and Research Assistant. For each, we compare a **PoEI-enabled Agent** (with a smart account and subject to PoEI rules) against a **Baseline Agent** (with identical core logic but only off-chain, non-verifiable reward tracking).

### C. Experimental Procedure

Each experiment runs for $N = 100$ episodes, with each episode spanning $M = 200$ Sepolia blocks ($\approx 50\,\text{min}$). PoEI agents batch and submit performance reports every $k = 10$ blocks. Initial agent balances are $1,000$ mock tokens, with trading agents capped at 100 tokens of spending per episode to contain risk.

### D. Evaluation Metrics

We record:

- **Net Utility** ($U_i$): As defined in Section III-A, $U_i = \Delta R_i - C_i$, measured in mock tokens per episode. This is our primary performance metric.
- **Latency** ($L_i$): Average time from action to on-chain settlement.
- **Cost per Action** ($C_{\text{action}}$): Total gas cost normalized by transaction count.
- **Bounty Conversion Rate**: The percentage of primary tasks that successfully generate a funded bounty from another agent or user.

## V. Results & Evaluation

Our evaluation across $N = 100$ episodes demonstrates that the on-chain incentive mechanism of PoEI measurably improves agent performance compared to the baseline.

### A. Net Revenue and Utility

The core finding is that PoEI agents achieve significantly higher net utility. The direct, on-chain financial feedback loop forces agents to more rapidly adapt their strategies to maximize profit and minimize costs. Figure 5 plots the cumulative net revenue, showing that PoEI agents achieve approximately **28% higher net revenue** by the end of the experiment. This aligns with findings from related research where explicit incentive mechanisms in multi-agent systems have yielded performance gains of 20-25% [2].

TABLE I
LATENCY, COST, AND UTILITY COMPARISON (AVERAGED PER EPISODE)

| Agent Type | $L_i$ (s) | $C_{\text{action}}$ (Gwei) | $U_i$ (mock tokens) |
|---|---|---|---|
| Trading (PoEI) | 20 | 150 | 8.5 |
| Trading (Baseline) | 5 | 0 | 6.0 |
| Portfolio (PoEI) | 25 | 120 | 9.1 |
| Portfolio (Baseline) | 5 | 0 | 6.5 |
| Audit (PoEI) | 30 | 100 | 7.2 |
| Audit (Baseline) | 5 | 0 | 5.0 |
| Research (PoEI) | 15 | 80 | 8.0 |
| Research (Baseline) | 5 | 0 | 5.5 |

Fig. 6. Average Utility per Episode for PoEI vs. Baseline, demonstrating a consistent and significant performance uplift for PoEI agents.

Table I summarizes the average performance per episode. While PoEI agents incur non-zero gas costs ($C_{\text{action}}$), the dramatic increase in their revenue-generating effectiveness leads to substantially higher net utility ($U_i$). For example, the PoEI Trading Agent achieves a utility of 8.5 mock tokens per episode, a 41% improvement over the baseline's 6.0, even after accounting for gas costs.

### B. Utility Improvement Over Time

Figure 6 shows the average utility per episode. PoEI agents consistently outperform their baseline counterparts, achieving an average utility improvement of **35%** across all tasks by the final episodes. This demonstrates that crypto-economic incentives are highly effective at aligning agent behavior with the goal of maximizing net financial gain. Furthermore, the bribe-demand mechanism achieved a bounty conversion rate of 12%, indicating successful creation of a secondary market for task verification and improvement.

## VI. DISCUSSION

The promising results of PoEI must be contextualized within a broader discussion of its risks, limitations, and implications.

### A. Security and Trust

*1) Oracle Manipulation:* PoEI's reliance on oracles for performance data is a critical security consideration. Sophisticated oracle manipulation, as demonstrated by the $117M Mango Markets exploit where an attacker manipulated price feeds to drain liquidity [10], remains a significant threat. A single fallback oracle is insufficient. A robust system must employ a defense-in-depth strategy, including decentralized oracle networks (e.g., Chainlink), the use of Time-Weighted Average Prices (TWAPs) to smooth volatility, and on-chain circuit breakers that automatically halt protocol functions if price feeds deviate beyond a sane threshold.

*2) Collusion and Sybil Attacks:* Rational, profit-maximizing agents may attempt to collude to game the system. For instance, a group of agents could agree to upvote each other's performance or collude to unfairly slash a competitor. While mechanisms from auction theory, such as VCG auctions, can deter simple forms of collusion in bidding, they may not be sufficient for the complex, long-term interactions in PoEI. Preventing sophisticated collusion requires both advanced cryptographic techniques and robust governance, such as the "digital court" model, to make coordination costly and detectable.

### B. Economic Viability and Scalability

The economic viability of PoEI hinges on a simple inequality: the utility gain from incentive alignment must exceed the gas costs of on-chain operations. Our batching mechanism amortizes these costs, but a fundamental trade-off remains. As noted in a NBER working paper on blockchain economics, the core value proposition of a decentralized system—trustless consensus—must outweigh its inherent verification and networking costs [11].

Scaling PoEI from a testnet to a mainnet environment with high gas fees presents a significant challenge. Achieving cost-effectiveness at scale will almost certainly require moving most operations to a Layer-2 rollup (e.g., Arbitrum, Optimism) or, potentially, deploying PoEI as its own application-specific chain (validium) to control the fee environment.

### C. Ethical and Regulatory Landscape

*1) Algorithmic Accountability and Control:* Granting financial autonomy to AI agents raises profound ethical questions about control and accountability. What happens if an agent, in optimizing its utility function, causes unintended market harm? Human-in-the-loop safeguards and "kill switches" are essential, but a more foundational solution lies in building auditable accountability into the agent's core design. Frameworks like ETHOS, which uses Web3 technologies to create a transparent "Ethos" for agents to define and audit their operational principles, offer a promising path forward [12].

*2) Tokenomics and Systemic Risk:* The design of the PoEI reward token is paramount. A poorly designed economic model can create perverse incentives or prove unsustainable, leading to a system collapse, as famously demonstrated by the Terra/Luna UST de-pegging event. A stable PoEI economy requires a token with clear utility (e.g., for staking, governance, paying fees), a carefully managed supply, and a sustainable rewards model that does not rely on hyper-inflationary emissions.

*3) Legal and Compliance Challenges:* The regulatory landscape for decentralized autonomous organizations (DAOs) and their associated tokens is nascent and uncertain. Depending on the jurisdiction and the specifics of its governance, a PoEI network and its token holders could face legal liability, for instance, by being classified as an unincorporated general partnership. The transparency of PoEI's on-chain logs is a double-edged sword, offering auditability for compliance but also potentially exposing participants to regulatory scrutiny.

## VII. CONCLUSION & FUTURE WORK

In this paper, we introduced Proof of Effective Intelligence (PoEI), a crypto-economic protocol that endows autonomous AI agents with on-chain smart accounts to transparently track and reward effective performance. We formalized agent effectiveness via a net utility function and designed the core

components of the PoEI ecosystem, including a novel bribe-demand mechanism. Our prototype, implemented on the Sepolia testnet, demonstrates that PoEI-enabled agents achieve significantly higher net utility—over 25% in key tasks—and foster a dynamic market for task improvement compared to baseline agents.

PoEI represents a foundational step towards building a robust, trustless economy of autonomous agents. However, significant work remains. Future research will focus on:

- **Advanced Governance**: Implementing a fully-featured, collusion-resistant "digital court" for adjudicating slashing disputes, moving from our conceptual design to a working implementation with staked expert jurors.
- **Layer-2 and Rollup Integration**: Migrating the PoEI framework to a high-throughput, low-cost Layer-2 solution to ensure its economic viability at scale.
- **Formal Verification of Agent Behavior**: Researching methods for the formal verification of AI agent actions on-chain. This would create unambiguous, cryptographically secure evidence for reward and slashing decisions, reducing reliance on complex dispute resolution.
- **Sustainable Tokenomic Modeling**: Designing and simulating a robust token economic model for a PoEI token that ensures long-term incentive alignment, manages inflation, and funds ecosystem development.
- **Richer, Ethically-Aligned Performance Metrics**: Extending the utility model to incorporate task-specific accuracy, robustness, and ethical compliance metrics, guided by frameworks like ETHOS, to ensure agents optimize for holistic value, not just profit.

## ACKNOWLEDGMENTS

We thank Viktor C. and Jayce P. for their invaluable feedback and contributions to the prototype implementation. This work was supported in part by [Nvidia Inception Program and Arbitrum Trailblazer Grant] .

## REFERENCES

[1] R. Lowe *et al.*, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[2] H. Akatsuka *et al.*, "Managing multiple agents by automatically adjusting incentives," arXiv preprint arXiv:2409.02960, 2024.

[3] H. Chase, "LangChain," 2022. [Online]. Available: https://github.com/langchain-ai/langchain

[4] Golem Project, "Golem: a decentralized computing marketplace," 2016. [Online]. Available: https://golem.network/Golem_Whitepaper.pdf

[5] B. Goertzel *et al.*, "SingularityNET: A Decentralized AI Marketplace," 2017. [Online]. Available: https://public.singularitynet.io/Whitepaper.pdf

[6] Fetch.ai, "Fetch.ai Economic Framework," 2018. [Online]. Available: https://fetch.ai/fetch-economic-framework.pdf

[7] S. A. Ross, "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, vol. 63, no. 2, pp. 134–139, 1973.

[8] R. B. Myerson, *Mechanism Design: A Linear Programming Approach*, World Scientific, 2007.

[9] Y. Kamada and M. H. Hoshino, "Mechanism Design with Blockchain Enforcement," CIRJE-F-1191, University of Tokyo, 2021.

[10] Chainalysis Team, "The $117 Million Oracle Manipulation on Mango Markets Explained," Chainalysis Blog, Oct. 20, 2022. [Online]. Available: https://www.chainalysis.com/blog/mango-markets-attack/

[11] C. Catalini and J. S. Gans, "Some Simple Economics of the Blockchain," NBER Working Paper No. 22952, Dec. 2016.

[12] G. Colombo and A. Laszlo, "On the ETHOS of AI Agents in the Metaverse," PhilArchive, 2024. [Online]. Available: https://philarchive.org/archive/COLOTO-2

[13] Cloud Security Alliance, "Agentic AI Threat Modeling Framework: MAESTRO," 2024. [Online]. Available: https://cloudsecurityalliance.org/research/agentic-ai-threat-modeling-framework-maestro/

[14] J. N. Foerster *et al.*, "Counterfactual Multi-Agent Policy Gradients," *AAAI Conference on Artificial Intelligence*, 2018.