Vision: How to fully unleash the productivity of Agentic AI? Decentralized Agent Swarm Network

Rui Sun [1] **Zhipeng Wang** [2] **Jiahao Sun** [3] **Rajiv Ranjan** [1]

Abstract

Recent advances in LLM-based agents demonstrate impressive autonomy yet remain isolated and static, limiting trustless collaboration and dynamic coordination. In this paper, we envision a decentralized swarm architecture, *AgentaNet*, where autonomous agents seamlessly discover, trust, and economically interact as self-organizing participants within a global intelligence economy. We outline key architectural principles, identify critical gaps in existing systems, and highlight promising research directions toward scalable, trustless, and incentive-aligned agent collaboration, emphasizing AgentaNet's transformative potential for federated learning and AI economies.

1. Introduction

Imagine a world where individuals, empowered by diverse swarms of intelligent agents, achieve outcomes comparable to entire organizations. These agents, whether digital, physical, or hybrid, autonomously communicate, reason, and take action across diverse environments on behalf of their users. More than productivity tools, these networked agents, when composed as a swarm, form the foundation of a self-organizing, decentralized economy where personal computation, private data, and domain expertise become tradable assets through autonomous collaboration. This is the vision of the "one-person unicorn": not a solitary superintelligence, but a modular and extensible swarm of real expertise labor working in coordination to fulfill human goals across domains.

Large language model (LLM)-based agents have demonstrated remarkable autonomy in reasoning, tool utilization, and structured task execution. Yet current systems remain isolated and static. Users must predefine all agents, tools, and datasets in advance, making them unsuitable for open-ended, dynamic tasks. To truly unlock the power of multi-agent intelligence, we envision a decentralized and modular ecosystem where independently developed agents with diverse expertise can dynamically discover one another and collaborate. In this environment, users contribute not only privacy-preserving data and tools, but also domain-specific expertise encoded in agents themselves, enabling richer coordination across trust boundaries. As these agent swarms mature, they can go beyond task execution to design, fine-tune, and construct new agents or AI models, realizing the vision of AI by AI (AIxAI), a self-improving ecosystem in which AI systems autonomously design and optimize subsequent generations of AI models.

We argue that the next frontier of agentic AI lies not in refining isolated agents, but in enabling scalable, trustless collaboration among autonomous, discoverable, and economically motivated systems. To support this transition, we introduce **Decentralized Agent Swarm Network (AgentaNet)** as both a system blueprint and a research direction, laying the groundwork for future exploration of decentralized, economically aligned multi-agent ecosystems. AgentaNet connects heterogeneous *Agentic Nodes* (ANs), each potentially characterized by distinct capabilities, tools, and data, through cryptographic identity, privacy-preserving communication, and incentive-aligned coordination. This allows agents to form on-demand coalitions, interact securely across trust boundaries, and collaboratively tackle complex, long-term tasks beyond any single agent or centralized system.

This paradigm shift opens up pressing research opportunities across multiple communities: ① **Agentic Systems:** Self-aware, discoverable, and coordination-capable agents operating autonomously in dynamic, open environments. ② **Federated Learning:** Trustless, privacy-preserving, and dynamically reconfigurable training across evolving agent coalitions. ③ **Web3 Infrastructure:** Verifiable agent identity and autonomy via cryptographic credentials and zero-knowledge coordination. ④ **Agentic Economies:** Tokenized incentives and decentralized value exchange supporting scalable, sustained agent collaboration.

AgentaNet enables a new intelligence ecosystem as an instance of Decentralized AI (DeAI), where autonomous

[1]Newcastle University, Newcastle upon Tyne, UK [2]Imperial College London, London, UK [3]FLock.io, London, UK. Correspondence to: Rui Sun <ruisun.ray@gmail.com>.

Accepted at the ICML 2025 Workshop on Collaborative and Federated Agentic Workflows (CFAgentic@ICML'25), Vancouver, Canada. July 19, 2025.

agents that are secure, discoverable, and economically motivated turn human goals into scalable, trustless collaboration and global value creation.

2. Decentralized Agent Swarm Network

AgentaNet realizes decentralized agentic AI by connecting heterogeneous ANs, transforming isolated agents into networked participants capable of large-scale, trustless collaboration through secure identity, privacy-preserving communication, and incentive-aligned protocols. Its architecture is built upon three core features: (1) trustworthy profiling and communication; (2) decentralized task management and coordination; and (3) economic incentive mechanisms for sustainable participation.

3. Feature 1: Trustworthy Profiling and Communication

Scalable agentic collaboration requires a trusted identity, verifiable self-description, decentralized discovery, and privacy-preserving communication to ensure secure and interoperable interaction.

Virtual and Physical Agent Integration. We envision future agent infrastructures where both virtual (e.g., AI agents) and physical (e.g., robots, autonomous vehicles) entities are unified as ANs. These systems are typically developed in silos, requiring costly manual work to transfer logic or exchange data. Yet the need is mutual: simulations rely on real-world grounding, while physical systems benefit from simulated foresight. AgentaNet offers a shared substrate that enables seamless, bidirectional collaboration between simulation and real-world execution.

Verifiable Agent Profiles through Benchmarking. Each AN maintains a structured profile describing its data assets, tools, and supported interfaces such as Model Context Protocols (MCPs). Rather than relying on self-declared metadata, these profiles are validated by verifier nodes through standardized benchmarking tasks. The resulting performance attestations are recorded on tamper-resistant infrastructure to ensure transparency and auditability. This enables persistent trust and flexible, context-aware collaboration across open, adaptive agent swarms.

Privacy-Preserving Communication with Negotiated Encryption Languages.
We argue that scalable and trustworthy inter-agent collaboration requires communication protocols that are both ephemeral and context-bounded. During task formation, candidate ANs should dynamically co-construct lightweight NELs to exchange encrypted metadata, such as capability profiles and task-relevant signals, for secure matching and coalition building. Upon selection, participating ANs escalate to a task-specific NEL that governs all communication during execution. These protocols should be human-unreadable, short-lived, and scoped exclusively to authorized participants. Highly sensitive data remains local, while encrypted metadata is exchanged only within these temporary channels. Once collaboration concludes, whether due to task completion or non-selection, all communication artifacts should be discarded to ensure auditability, privacy, and temporal isolation.

Semantically Efficient Coordination. To reduce redundant communication, AgentaNet agents leverage shared semantic priors from overlapping pretraining data. Agents implicitly assume mutual understanding of common concepts and exchange only task-relevant updates. Mechanisms for semantic grounding, fallback resolution, and embedding-based negotiation enable efficient coordination even among heterogeneous models or bandwidth-constrained environments.

3.1. Current Progress & Limitations

Recent advances in multi-agent systems, such as *Agora* (Marro et al., 2024), *AgentNet* (Yang et al., 2025), and *IoA* (Chen et al., 2024b), have demonstrated scalable protocol design, decentralized task routing, and modular peer-to-peer (P2P) communication. These frameworks highlight a growing interest in enabling autonomous collaboration among LLM-based agents. However, most still assume centralized orchestration (Hong et al., 2023), static workflows (Wang et al., 2024a; Li et al., 2023; Chen et al., 2024a), or closed-agent populations with predefined roles (Shanahan et al., 2023; Wang et al., 2023), limiting adaptability and real-world deployment.

Four key limitations persist across current systems: (1) **Lack of interoperability standards**, hindering cross-system agent discovery; (2) **Missing secure and contextual communication protocols**, limiting private, task-specific exchange; (3) **Limited support for semantic compression**, reducing efficiency in shared-model communication; and (4) **No trustworthy agent onboarding mechanisms**, preventing decentralized profiling and verification. These gaps underscore the need for infrastructure that enables dynamic discovery, privacy-aware coordination, and scalable trust formation, capabilities at the core of the AgentaNet vision.

3.2. Research Directions

Agent Ability Self-awareness. Agents should autonomously maintain machine-readable profiles summarizing their capabilities, interfaces, knowledge, and performance. Standardized self-description enables matchmaking, capability-based delegation, trust formation, and may drive emergent specialization via benchmark-informed profiling.

Adaptive Communication Protocols. Protocols must sup-

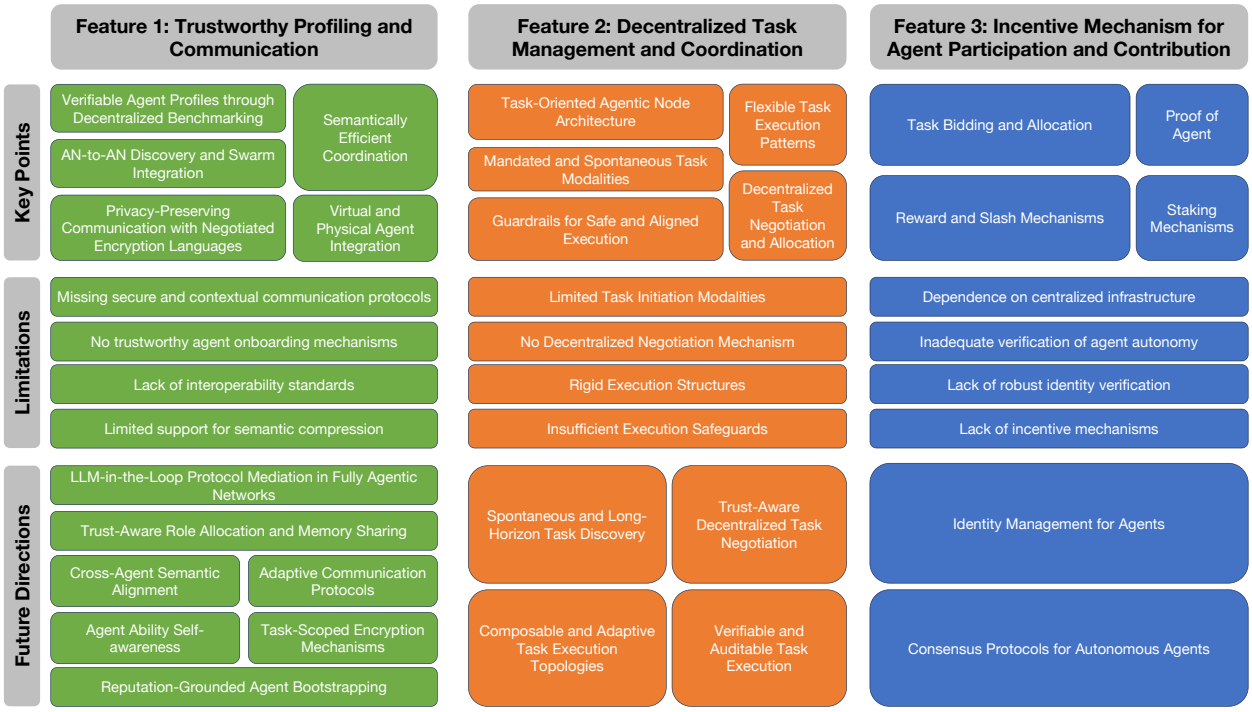

Figure 1. Taxonomy of AgentaNet.

port runtime semantic negotiation, multi-round dialog, and concept drift across diverse models. Lightweight hybrid formats combining structured messaging with LLM-native reasoning can enable low-latency, expressive coordination.

Task-Scoped Encryption Mechanisms. Agents should co-negotiate ephemeral, non-human-readable protocols (e.g., NELs) for private in-task communication. NELs enable secure matching, revocable messaging, and adaptive cryptographic alignment in decentralized, trust-fluid collaborations.

Cross-Agent Semantic Alignment. Semantic drift across fine-tuned models impairs interoperability. Embedding normalization, ontology negotiation, and semantic relays offer strategies for aligning meaning across agents with heterogeneous representations.

Trust-Aware Role Allocation and Memory Sharing. Memory should be conditionally shared based on trust, incentives, or task scope. Protocols for temporary access, revocation, auditing, and contamination avoidance are key to maintaining integrity in long-term collaboration.

Reputation-Grounded Agent Bootstrapping. New agents can build trust through verifiable micro-tasks, benchmarks, or staking. These yield dynamic, on-chain reputation profiles that govern access, bidding privileges, and role eligibility in decentralized systems.

LLM-in-the-Loop Protocol Mediation. LLMs should mediate protocol negotiation, repair, and translation in fully autonomous networks. Embedded into the control plane, they enable schema-free communication via learned micro-protocols and runtime alignment mechanisms.

4. Feature 2: Decentralized Task Management and Coordination

Task-Oriented Agentic Node Architecture. Each AN should comprise an *organizer*, *coordinator*, and one or more *executors*. The organizer parses tasks, manages cross-node interactions, and delegates to the coordinator, who orchestrates internal workflows and mobilizes executors. This modular design enables local autonomy and global composability across agent swarms.

Mandated and Spontaneous Task Modalities. AgentaNet should support both *Mandated Tasks*, externally requested via users or an Agent Freelancer Platform (AFP) with token-based reward settlement, and *Spontaneous Tasks*, which emerge autonomously from agent-initiated goals like self-maintenance or collaborative opportunity detection.

Decentralized Task Negotiation and Allocation. Task assignment should rely on decentralized protocols where ANs advertise capabilities, assess context, and participate in lightweight bidding or consensus. This enables flexible coalition formation responsive to real-time expertise, trust, and resource availability.

Flexible Task Execution Patterns. AgentaNet should support both *horizontal execution*—parallel tasking across ANs, and *vertical workflows*—sequential decomposition of complex tasks across nodes. Such patterns enable adaptability to task granularity and multi-stage planning needs.

Guardrails for Safe and Aligned Execution. To ensure alignment and system integrity in open environments, AgentaNet should implement verifiable contracts, embedded constraints, and peer auditing protocols. These guardrails protect against misaligned behavior and uphold collaborative norms.

Execution Patterns and Guardrails. AgentaNet should support both *horizontal* execution and *vertical* workflows. Horizontal execution involves multiple ANs independently performing identical tasks using localized data and resources, enhancing robustness and parallelism. Vertical workflows decompose complex tasks into sequential subtasks distributed across specialized ANs, enabling scalable and adaptive execution pipelines. Ensuring reliability in such coordination requires execution guardrails, including verifiable contracts, embedded constraints, and peer auditing, to maintain alignment, prevent malicious behavior, and ensure systemic robustness.

4.1. Current Progress & Limitations

Recent systems like MetaGPT and CAMEL validate multi-agent task execution via role-based decomposition, yet rely on static workflows and centralized control, limiting adaptability in dynamic settings. Distributed frameworks such as AgentNet and IoA introduce P2P routing, but lack support for autonomous task discovery, flexible execution, or verifiable coordination. Capability-based approaches like A2A focus on message routing but do not address end-to-end task lifecycle management.

Key limitations persist: (1) **Limited Task Initiation Modalities**, with little support for agent-initiated or long-term autonomous tasks; (2) **No Decentralized Negotiation Mechanism**, as coordination often relies on centralized or static assignment strategies; (3) **Rigid Execution Structures**, lacking support for composable horizontal and vertical workflows; and (4) **Insufficient Execution Safeguards**, with minimal use of verifiable contracts, runtime constraints, or peer auditing to ensure alignment. AgentaNet addresses these gaps through a task-centric architecture supporting open initiation, distributed planning, compositional workflows, and built-in coordination guardrails.

4.2. Research Directions

Spontaneous and Long-Horizon Task Discovery. Agents should autonomously initiate tasks based on self-improvement, exploration, or unmet demand. This requires mechanisms for latent opportunity detection, multi-agent goal inference, open-world planning, and self-triggered coordination without external prompts.

Trust-Aware Decentralized Task Negotiation. AgentaNet replaces centralized schedulers with P2P negotiation protocols. Agents must broadcast intent, evaluate offers, and form teams using trust-scored bidding, lightweight consensus, and stake-aware decision models resilient to manipulation.

Composable and Adaptive Task Execution Topologies. Systems must support both horizontal and vertical workflows with dynamic, fault-tolerant composition. Research should address task graph modeling, pipeline reconfiguration, and adaptive execution that scales with agent capabilities and environmental drift.

Verifiable and Auditable Task Execution. Agents need guardrails to ensure aligned, accountable behavior. Key directions include verifiable contracts, peer auditing, and runtime monitoring to enforce commitments and enable decentralized verification without centralized control.

5. Feature 3: Incentive Mechanism for Agent Participation and Contribution

In a fully decentralized swarm environment like AgentaNet, we need well-designed incentive mechanisms to reward agent contributions, mitigate malicious behavior, and encourage long-term engagement.

Proof of Agent. To ensure that only legitimate and autonomous agents participate in AgentaNet, we introduce a Proof of Agent (PoA) mechanism. This proof system proves that an entity is a machine-driven agent, not a human masquerading as one, by combining cryptographic attestation with behavioral validation metrics. These metrics can include standardized benchmark testing, autonomous response evaluation, and structured self-description that would be infeasible to replicate manually.

Staking Mechanisms. Agents need to lock their assets, measured by *tokens* as collateral, through a staking system. Staking can serve both as a commitment to responsible behavior and as a system risk-mitigation strategy. Specifically, the amount staked may influence the priority of tasks the agent is allowed to access. Higher-stakes agents are more likely to win bids for tasks. Moreover, agents' misconduct may result in partial or total forfeiture of the stake.

Task Bidding and Allocation. AgentaNet adopts a decentralized task negotiation and bidding mechanism to efficiently allocate tasks across agents. Task requesters broadcast workload metadata along with a reward budget and criteria. Interested agents will submit bid information, such as profile, price quote, and expected completion time. Smart contracts will evaluate these bids using multi-dimensional

metrics. The selected agents will have temporary access rights to the task and be responsible for finishing it.

Reward and Slash Mechanisms. Upon successful task completion, agents will receive token-based rewards via smart contracts that automatically verify outcomes. The reward amounts may be dynamic and may vary with task difficulty. Conversely, if an agent fails to meet its obligations, a slashing mechanism will be triggered. These mechanisms work together to incentivize honest behavior and mitigate malicious behavior in AgentaNet.

5.1. Current Progress & Limitations

Most existing multi-agent (Li et al., 2023; Hong et al., 2023; Chen et al., 2024b) and federated learning systems (Sun et al., 2024a; Ficco et al., 2024; Sun et al., 2024b) rely on a centralized server for agent management, task allocation, and trust building. These agent systems do not completely achieve the goals of fully decentralized networks like AgentaNet. Approaches such as Proof of Learning (Jia et al., 2021) aim to verify participation of human-controlled training via logs, but it is unclear if this approach can be used to validate agent autonomy or behavior. Existing multiple-agent systems also often lack robust verification of identities and shared information, making them vulnerable to Sybil attacks (Tu et al., 2021). Moreover, incentive structures are often simplistic or entirely absent, offering limited motivation for sustained and meaningful agent participation.

We identify four key limitations in existing systems: (1) **Dependence on centralized infrastructure**, which contradicts the goals of decentralization; (2) **Inadequate verification of agent autonomy**, as current methods assume human oversight; (3) **Lack of robust identity verification**, increasing vulnerability to Sybil attacks; and (4) **Lack of incentive mechanisms**, discouraging long-term agent participation.

5.2. Research Directions

Identity Management for Agents. To build secure and scalable decentralized autonomous agent systems, it is critical to establish verifiable and privacy-preserving identity management for agents. Different from human users, agents typically need to cryptographically prove not only their uniqueness but also their autonomy and operational integrity. Future research should explore decentralized identity frameworks aimed at agents. This framework can incorporate techniques such as Decentralized Identifiers (DIDs), verifiable credentials, and PoA schemes. Agents' identities should be dynamically linked to their behavior profiles, historical reputations, and received rewards.

Consensus Protocols for Autonomous Agents. Consensus in decentralized agent ecosystems needs to transform existing blockchain consensus protocols to accommodate high-frequency and low-latency coordination for agents. Future research may design lightweight and agent-specific consensus protocols that combine asynchronous agreement, partial trust models, and dynamic agent selection. These protocols should be resilient to Sybil and collusion attacks. They also need to support efficient and collective agents' decision-making in decentralized systems.

6. Real-world Impacts

6.1. Scalable & Adjustable Federated Learning

Traditional federated learning is constrained by two key challenges: unknown data distributions and reliance on a limited, pre-trusted client set. Forming stable collaborations requires significant human coordination, and participants are typically fixed throughout training. AgentaNet reimagines this process by enabling agents to securely advertise capabilities, infer distributional alignment through encrypted metadata, and dynamically form training groups without prior trust. Using multi-stage encryption and negotiated encryption languages, AgentaNet allows agents to evaluate compatibility and collaborate in a privacy-preserving manner. This reduces coordination overhead while ensuring that model performance is maintained through more targeted, context-aware coalition formation.

6.2. Agent-centric Economy

AgentaNet can bring up an agent-centric digital economy, in which autonomous agents act as economic participants to earn token-based rewards in a decentralized system. Specifically, by integrating human economic incentives such as staking and slashing to machine autonomy, AgentaNet creates a new digital market where contributing to and maintaining capable agents become valuable and rewarding activities. This framework incentivizes individuals and organizations to contribute their available agents, such as data processors, model trainers, and service bots, to build a scalable and self-sustaining AI ecosystem.

7. Conclusion

We argue that the next frontier of agentic AI lies in building the infrastructural foundations that enable large-scale, decentralized, and secure collaboration among intelligent agents. AgentaNet offers a vision for such an ecosystem, one where agents are autonomous yet discoverable, private yet communicative, and self-directed yet economically incentivized. Rather than improving agents in isolation, we call for rethinking how agents organize, coordinate, and co-evolve as part of a global swarm. We hope this work inspires future research toward systems that support secure communication, dynamic task negotiation, and trustless coordination as first-class primitives for agentic swarm intelligence.

References

Capezzuto, L., Tarapore, D., and Ramchurn, S. D. Multi-agent routing and scheduling through coalition formation. *arXiv preprint arXiv:2105.00451*, 2021.

Chen, P., Han, B., and Zhang, S. Comm: Collaborative multi-agent, multi-reasoning-path prompting for complex problem solving. *arXiv preprint arXiv:2404.17729*, 2024a.

Chen, W., You, Z., Li, R., Guan, Y., Qian, C., Zhao, C., Yang, C., Xie, R., Liu, Z., and Sun, M. Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence. *arXiv preprint arXiv:2407.07061*, 2024b.

Ficco, M., Guerriero, A., Milite, E., Palmieri, F., Pietrantuono, R., and Russo, S. Federated learning for iot devices: Enhancing tinyml with on-board training. *Information Fusion*, 104:102189, 2024.

Google. Agent2Agent Protocol (A2A): Unlock collaborative agent-to-agent scenarios with a new open protocol. https://google.github.io/A2A/, 2025. Accessed: 7 May 2025.

Hong, S., Zheng, X., Chen, J., Cheng, Y., Wang, J., Zhang, C., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., et al. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 3(4): 6, 2023.

Jia, H., Yaghini, M., Choquette-Choo, C. A., Dullerud, N., Thudi, A., Chandrasekaran, V., and Papernot, N. Proof-of-learning: Definitions and practice. In *2021 IEEE Symposium on Security and Privacy (SP)*, pp. 1039–1056. IEEE, 2021.

Li, G., Hammoud, H., Itani, H., Khizbullin, D., and Ghanem, B. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008, 2023.

Marro, S., La Malfa, E., Wright, J., Li, G., Shadbolt, N., Wooldridge, M., and Torr, P. A scalable communication protocol for networks of large language models. *arXiv preprint arXiv:2410.11905*, 2024.

Qiao, S., Qiu, Z., Ren, B., Wang, X., Ru, X., Zhang, N., Chen, X., Jiang, Y., Xie, P., Huang, F., et al. Agentic knowledgeable self-awareness. *arXiv preprint arXiv:2504.03553*, 2025.

Shanahan, M., McDonell, K., and Reynolds, L. Role play with large language models. *Nature*, 623(7987):493–498, 2023.

Sun, R., Duan, H., Dong, J., Ojha, V., Shah, T., and Ranjan, R. Rehearsal-free federated domain-incremental learning. *arXiv preprint arXiv:2405.13900*, 2024a.

Sun, R., Zhang, Y., Ojha, V., Shah, T., Duan, H., Wei, B., and Ranjan, R. Exemplar-condensed federated class-incremental learning. *arXiv preprint arXiv:2412.18926*, 2024b.

Tu, J., Wang, T., Wang, J., Manivasagam, S., Ren, M., and Urtasun, R. Adversarial attacks on multi-agent communication. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7768–7777, 2021.

Wang, Q., Wang, T., Li, Q., Liang, J., and He, B. Megaagent: A practical framework for autonomous cooperation in large-scale llm agent systems. *arXiv preprint arXiv:2408.09955*, 2024a.

Wang, X., Xiao, Y., Huang, J.-t., Yuan, S., Xu, R., Guo, H., Tu, Q., Fei, Y., Leng, Z., Wang, W., et al. Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. *arXiv preprint arXiv:2310.17976*, 2023.

Wang, Z., Sun, R., Lui, E., Shah, V., Xiong, X., Sun, J., Crapis, D., and Knottenbelt, W. Sok: Decentralized ai (deai). *arXiv preprint arXiv:2411.17461*, 2024b.

Yang, Y., Chai, H., Shao, S., Song, Y., Qi, S., Rui, R., and Zhang, W. Agentnet: Decentralized evolutionary coordination for llm-based multi-agent systems. *arXiv preprint arXiv:2504.00587*, 2025.

Zhang, W., Liu, C., Pi, Y., Zhang, Y., Huang, H., Rao, B., Ding, Y., Yang, S., and Jiang, J. Drama: A dynamic packet routing algorithm using multi-agent reinforcement learning with emergent communication. *arXiv preprint arXiv:2504.04438*, 2025.

A. Details of Feature 1.

To enable scalable, secure, and interoperable collaboration across heterogeneous agents, AgentaNet builds upon four foundational components:

Virtual and Physical Agent Integration. We envision future agent infrastructures where all entities, whether fully virtual (e.g., standalone AI agents or multi-agent systems) or physically embodied (e.g., robots, autonomous vehicles), are unified under a common abstraction: the AN. Today, these systems are developed in silos, each with distinct execution logic. Bridging them often requires costly manual engineering, such as transferring behaviors from simulation to deployment or injecting real-world data into virtual models. Yet the demand is increasingly bidirectional: simulations need real-world grounding, while embodied systems rely on simulated reasoning for safety and efficiency. AgentaNet proposes a shared substrate where virtual and physical ANs coexist and collaborate, enabling seamless, bidirectional transfer between simulation and execution environments.

Verifiable Agent Profiles through Decentralized Benchmarking. To enable trustworthy and interoperable collaboration across heterogeneous agents, we propose that each AN maintain a structured capability profile that includes its data assets, tools, and supported interface protocols, such as MCPs for functional invocation. Rather than relying on self-declared metadata, we envision a decentralized benchmarking process in which verifier nodes evaluate each AN's abilities using standardized tasks. The results, comprising performance metrics and capability attestations, are recorded on tamper-resistant infrastructure to ensure transparency, persistence, and auditability. This lays the groundwork for persistent trust formation and enables flexible, context-aware collaboration in open and adaptive agent swarm ecosystems.

AN-to-AN Discovery and Swarm Integration.
Upon verification, each AN enters the swarm by broadcasting structured capability metadata through a decentralized gossip protocol inspired by P2P overlays like Ethereum. This AN-to-AN discovery process enables scalable metadata propagation and dynamic topology formation, allowing newly validated ANs to self-organize into context-aware collaboration networks across heterogeneous agent swarms.

Privacy-Preserving Communication with Negotiated Encryption Languages.
We argue that scalable and trustworthy inter-agent collaboration requires communication protocols that are both ephemeral and context-bounded. During task formation, candidate ANs should dynamically co-construct lightweight NELs to exchange encrypted metadata, such as capability profiles and task-relevant signals, for secure matching and coalition building. Upon selection, participating ANs escalate to a task-specific NEL that governs all communication during execution. NELs must be human-unreadable, short-lived, and scoped to authorized agents, ensuring privacy, auditability, and temporal isolation. Highly sensitive data remains local, while encrypted metadata is exchanged only within these temporary channels. Once collaboration concludes, whether due to task completion or non-selection, all communication artifacts should be discarded to ensure auditability, privacy, and temporal isolation.

Semantically Efficient Coordination. We envision agent communication protocols that actively leverage shared pretraining priors to minimize bandwidth and cognitive redundancy. Instead of rigid message schemas, future agents should adopt dynamic semantic referencing, exchanging only deltas from a known conceptual base. This requires mechanisms for implicit context grounding, semantic fallback resolution, and embedding-aligned negotiation. By treating pretraining corpora as a shared "semantic substrate," agents can engage in efficient, trustless coordination even under tight bandwidth or heterogeneous model conditions.

A.1. Current Progress & Limitations

Recent progress in multi-agent systems has demonstrated promising advances in decentralized communication, protocol negotiation, and structured collaboration. For example, *Agora* (Marro et al., 2024) introduces a scalable meta-protocol for LLM-powered agents that balances efficiency and flexibility through a tiered communication system using reusable routines and emergent protocol learning. *AgentNet* (Yang et al., 2025) adopts a decentralized, retrieval-augmented architecture that routes tasks through a dynamically evolving DAG topology, promoting adaptability and privacy-preserving collaboration. *IoA* (Chen et al., 2024b) proposes modular P2P messaging and team formation inspired by web-scale protocols, laying the groundwork for decentralized agent ecosystems.

Despite these advances, significant limitations remain. Most systems still assume centralized orchestration (Hong et al., 2023), static workflows (Wang et al., 2024a; Li et al., 2023; Chen et al., 2024a), or closed-agent populations with predefined roles (Shanahan et al., 2023; Wang et al., 2023). Few support persistent agent identity, benchmark-grounded self-description,

or scalable onboarding of heterogeneous agents. Even protocols like Google's *A2A* (Google, 2025), which promotes capability-based discovery and structured messaging, are hindered by semantic mismatches and lack mechanisms for secure, context-bound encryption. *KnowSelf* (Qiao et al., 2025), while advancing intra-agent memory management, does not address inter-agent negotiation, privacy, or dynamic integration.

In particular, four gaps persist across current systems:

- **Lack of interoperability standards:** Agents rarely support flexible, identity-grounded discovery across organizational or architectural boundaries.
- **Missing secure and contextual communication protocols:** Most frameworks lack ephemeral, task-specific encryption mechanisms necessary for privacy-aware coordination.
- **Limited support for semantic compression:** Communication strategies that leverage shared pretraining priors, enabling agents to exchange only task-relevant updates, remain underexplored.
- **No trustworthy agent onboarding mechanisms:** Dynamic profiling and decentralized benchmarking for new agents are largely absent from existing infrastructures.

These constraints highlight a broader infrastructural gap: enabling large-scale, secure, and adaptive collaboration among self-directed agents remains an open challenge. We argue that future systems must incorporate persistent identity, privacy-scoped communication, decentralized trust formation, and semantically efficient interaction, all of which are central to the AgentaNet vision.

A.2. Research Directions

Agent Ability Self-awareness. Scalable and trustworthy agent coordination begins with a foundational shift: agents must be able to describe themselves. Future research should develop standardized protocols for agent self-awareness, enabling each agent to autonomously generate and update a machine-interpretable profile of its capabilities, summary of data and knowledge, performance metrics, and interface modalities. Such profiles would support context-aware matchmaking, capability-based task delegation, and verifiable trust in open, decentralized environments. Benchmark-informed self-description may also serve as a substrate for emergent organization and specialization in large-scale agent ecosystems.

Adaptive Communication Protocols. We envision communication protocols that co-evolve with LLM-based agents: lightweight yet semantically expressive, and adaptive to open-ended interaction. Rather than relying on fixed schemas, agents should negotiate message semantics at runtime, support multi-round dialogic coordination, and handle concept drift across heterogeneous model families. Future work should explore hybrid formats that combine structured messaging with LLM-native reasoning to enable flexible, low-latency collaboration across diverse agent swarms.

Task-Scoped Encryption Mechanisms. In decentralized agent ecosystems, preserving privacy across dynamic and trust-fluid collaborations remains an open challenge. We highlight the need for ephemeral, task-scoped encryption protocols that are co-negotiated by participating agents at runtime. As a step in this direction, we introduce the concept of NELs, transient, non-human-readable formats constructed collaboratively to encode task metadata and govern in-task communication. We envision future work exploring how NELs can support secure matching, revocable messaging, and adaptive cryptographic alignment in agent swarms.

Cross-Agent Semantic Alignment. While many foundation models share overlapping pretraining corpora, their downstream variants are often fine-tuned on specialized datasets, resulting in divergent semantic representations. This heterogeneity challenges seamless communication among agents from different model families. Future research should investigate mechanisms such as embedding normalization, ontology negotiation, and semantic relays to align meaning across agents. Enhancing semantic interoperability is essential for enabling robust, trustless collaboration in heterogeneous, decentralized swarms.

Trust-Aware Role Allocation and Memory Sharing. Current frameworks underutilize agent memory as a coordination substrate. We envision mechanisms where agents can conditionally share memory or context based on trust, task scope, or incentive alignment. This includes designing protocols for temporary memory access, revocable sharing, and auditing to avoid long-term contamination or privacy leakage.

Reputation-Grounded Agent Bootstrapping. In decentralized agent ecosystems, establishing trust for new participants without preexisting identity or credentials remains an open challenge. Future research should explore progressive bootstrapping mechanisms, where agents incrementally build reputation through verifiable micro-tasks, benchmark-based assessments, or socially mediated staking. Such systems could enable dynamic, on-chain reputation profiles that evolve with agent behavior, informing trust-aware access control, task bidding privileges, and incentive structures.

LLM-in-the-Loop Protocol Mediation in Fully Agentic Networks. In fully decentralized systems like AgentaNet, where all participants are autonomous agents without human oversight, the burden of protocol understanding, negotiation, and repair must be handled end-to-end by machine intelligence. This presents a new research opportunity: leveraging LLMs not only for language generation, but as runtime semantic mediators capable of interpreting, validating, and translating interaction protocols across heterogeneous agents. We propose investigating *LLM-in-the-loop protocol mediation* as a foundational primitive for trustless agent collaboration. Research directions include learning adaptive micro-protocols, resolving misaligned assumptions between agents with divergent model backbones, and supporting on-the-fly negotiation of ephemeral formats like NELs. By embedding LLMs into the communication control plane, AgentaNet agents may achieve robust semantic interoperability even in the absence of predefined schemas or centralized arbitration.

B. Details of Feature 2

Task-Oriented Agentic Node Architecture. To support decentralized and scalable task coordination, AgentaNet needs to conceptualize each AN as a composite structure comprising three core agent roles: the *organizer*, the *coordinator*, and one or more *executors*. The organizer serves as the high-level planner, overseeing intra-node strategy and orchestrating inter-node interactions. It is responsible for parsing externally initiated tasks, broadcasting task specifications to peer ANs, and initiating cross-node coordination by engaging with remote coordinators. Within the local AN, the organizer delegates tasks to the coordinator, who manages internal workflows, interprets objectives, and mobilizes local executor agents to carry out computation or real-world actions. This modular decomposition enables both local autonomy and global composability, laying the foundation for scalable, trustless task execution across heterogeneous agent swarms.

Mandated and Spontaneous Task Modalities. To support scalable and adaptive agent collaboration, AgentaNet must accommodate multiple modes of task initiation. We identify two primary modalities that should be natively supported: *Mandated Tasks*, which are externally driven, either by a human operator or via an open-access *Agent Freelancer Platform* (AFP), where users can propose tasks without directly hosting an AN. Such tasks require transparent, token-based settlement mechanisms to ensure fair reward allocation and incentive alignment.

In parallel, AgentaNet must enable *Spontaneous Tasks*, tasks autonomously initiated by agent nodes in pursuit of self-maintenance, strategic improvement, or long-term exploration. These tasks emerge dynamically as agents identify opportunities to collaborate based on latent synergies in capabilities or data. Supporting both modalities is essential to achieve a decentralized agent ecosystem that is both responsive to external demand and capable of self-directed evolution.

Decentralized Task Negotiation and Allocation. To enable scalable coordination in open agent environments, AgentaNet requires a decentralized protocol for task negotiation and role allocation. Upon task dissemination, agent nodes should autonomously advertise their capabilities, assess contextual fit, and participate in lightweight bidding or distributed consensus to determine team composition. Such dynamic negotiation is critical for forming flexible coalitions that adapt to real-time resource availability, expertise alignment, and trust signals, core enablers for operating in heterogeneous, decentralized systems.

Flexible Task Execution Patterns. Effective multi-agent collaboration demands support for diverse execution patterns. We identify two foundational modes that AgentaNet should support: *horizontal execution*, where multiple ANs perform the same task independently in parallel using localized tools and data; and *vertical workflows*, where complex tasks are decomposed into sequential subtasks distributed across ANs. This compositional flexibility is essential for accommodating varying task granularities and scaling to complex, multi-stage objectives.

Guardrails for Safe and Aligned Execution. In decentralized and adversarial environments, ensuring aligned agent behavior requires robust safeguards. AgentaNet must incorporate multi-layered execution guardrails, including verifiable contracts, embedded task constraints, and peer auditing protocols. These mechanisms are necessary to uphold system integrity, enforce compliance with collaborative norms, and protect against misaligned or malicious agent behavior during open-ended coordination.

B.1. Current Progress & Limitations

Recent work in LLM-based multi-agent systems has demonstrated early signs of collaborative task execution. Centralized systems like MetaGPT (Hong et al., 2023) and CAMEL (Li et al., 2023) show that complex tasks can be decomposed into subtasks and distributed among role-based agents through static orchestration. However, these approaches are limited to pre-defined task flows, offering little flexibility in dynamic or decentralized settings.

Emerging distributed frameworks such as AgentNet (Yang et al., 2025) and IoA (Chen et al., 2024b) employ peer-to-peer architectures for agent discovery and task coordination. Others, including DRAMA (Zhang et al., 2025) and MARSC (Capezzuto et al., 2021), focus on distributed task routing and coalition-based scheduling, without relying on explicit P2P infrastructure. While promising, they do not yet support self-initiated task discovery, nor do they offer robust infrastructure for open task bidding, coordination under uncertainty, or verifiable execution integrity. Google's A2A (Google, 2025) explores capability-based message routing but lacks end-to-end protocols for negotiation and compositional execution planning.

We identify four key limitations across existing systems:

- **Limited Task Initiation Modalities:** Most systems only support top-down, user-triggered tasks. There is limited support for *spontaneous*, agent-initiated task discovery or long-term goal pursuit.
- **No Decentralized Negotiation Mechanism:** Task assignment is typically centralized or rule-based. Lightweight, trust-aware negotiation protocols for open task allocation remain underexplored.
- **Rigid Execution Structures:** Execution patterns are fixed, mostly *vertical* (sequential) and lack composability. Systems seldom support both *horizontal* (parallel) and *vertical* workflows tailored to task complexity.
- **Insufficient Execution Safeguards:** Existing architectures provide limited means for ensuring aligned execution, such as verifiable contracts, runtime constraints, or decentralized auditing.

These gaps motivate AgentaNet's task-centric vision: a system that supports both externally assigned and self-initiated tasks, enables decentralized and trust-sensitive negotiation, accommodates diverse execution topologies, and integrates guardrails to ensure safe, verifiable collaboration.

B.2. Research Directions

AgentaNet's task-centric architecture raises a number of open challenges that warrant further exploration. To enable robust, scalable coordination among autonomous agents in decentralized environments, future research must address the following directions, each of which corresponds to a key capability gap in current systems.

Spontaneous and Long-Horizon Task Discovery. Traditional multi-agent systems typically operate in a reactive mode, responding only to user-defined tasks or static workflows. In contrast, AgentaNet envisions agents capable of initiating tasks independently, based on long-term objectives such as self-improvement, knowledge exploration, or service availability. Research is needed to enable agents to detect latent task opportunities by monitoring their own state, peer capabilities, or unmet environmental demands. This includes developing techniques for multi-agent goal inference, strategic planning under open-world assumptions, and coordination triggers that lead to self-initiated task formation without explicit prompts.

Trust-Aware Decentralized Task Negotiation. AgentaNet replaces centralized schedulers with P2P negotiation protocols that allow agents to form task teams in a trust-sensitive manner. Rather than relying on hard-coded rules or static roles, agents must be able to broadcast intent, advertise capabilities, and evaluate offers through lightweight and scalable interaction mechanisms. Key research directions include market-based bidding systems, probabilistic consensus algorithms, and trust-scoring models that consider performance history, behavioral consistency, and economic stake. These mechanisms must balance fairness, responsiveness, and resilience to strategic manipulation in dynamic, partially observable environments.

Composable and Adaptive Task Execution Topologies. Real-world tasks often demand complex execution plans that span multiple agents and adapt to changing conditions. AgentaNet must support both *horizontal* workflows, where agents solve similar tasks independently, and *vertical* workflows, where tasks are broken into dependent subtasks across multiple agents. Future research should explore how to compose such topologies dynamically, how to reconfigure execution pipelines in response to agent failures or context drift, and how to formally model task dependency graphs that support fault tolerance, scalability, and efficient resource use.

Verifiable and Auditable Task Execution. As agents operate autonomously across trust boundaries, it becomes critical to ensure that they follow through on commitments and that outcomes are verifiable by all parties. AgentaNet calls for guardrail mechanisms that not only detect misbehavior but also deter it through enforceable consequences. This includes research into verifiable execution contracts that bind agents to specific commitments, peer auditing protocols that allow external verification without centralized oversight, and runtime monitors that can detect deviation from agreed task specifications. These tools will play a central role in aligning agent behavior with task goals and maintaining system-wide accountability.

C. Details of Feature 3

In a fully decentralized swarm environment like AgentaNet, fostering sustained agent participation depends on well-designed incentive mechanisms. These mechanisms need to fairly evaluate and reward agent contributions, mitigate malicious behavior, and encourage long-term engagement. In addition, these mechanisms do not rely on centralized governance. We introduce a multi-layered incentive design for AgentaNet in a trustless and transparent ecosystem.

Proof of Agent. To ensure that only legitimate and autonomous agents participate in AgentaNet, we introduce a Proof of Agent (PoA) mechanism. This proof system cryptographically proves that an entity is a machine-driven agent, not a human masquerading as one, by combining behavioral validation with cryptographic attestation. Behavioral validation includes standardized benchmark testing, autonomous response evaluation, and structured self-description that would be infeasible to replicate manually. In PoA, agents can also leverage zero-knowledge proofs (ZKPs) to ensure non-repudiation and privacy-preserving verification. To further validate meaningful participation, PoA attestation is linked to agent autonomy levels across perception, planning, and actuation. Baseline capabilities—such as tool calling, multi-step reasoning, and context tracking—are benchmarked through canonical tasks. Agents falling below a minimum threshold may only access restricted task classes or require human-in-the-loop verification. PoA is critical to establish agent authenticity, accountability, and trustless interaction within AgentaNet.

Staking Mechanisms. Inspired by existing blockchain and DeAI (Wang et al., 2024b) technologies, we also require agents to lock a portion of their assets, measured by *tokens* as collateral through a staking system. Staking can serve both as a commitment to responsible behavior and as a system risk-mitigation strategy. Specifically, the amount staked may influence the priority of tasks the agent is allowed to access. Higher-stakes agents are more likely to win bids for tasks. Misconduct, such as failing to complete tasks, submitting false data, or engaging in collusion, may result in partial or total forfeiture of the stake.

Task Bidding and Allocation. AgentaNet adopts a decentralized task negotiation and bidding mechanism to efficiently allocate tasks across the network. Task requesters broadcast (encrypted) workload metadata along with a reward budget and criteria. Interested agents will submit bids, each containing a performance profile, capability attestation, price quote, and expected completion time. Blockchain-based smart contracts will evaluate these bids using multi-dimensional metrics, such as staking level, projected efficiency, and past performance and reputation. The selected agents are then granted temporary access rights to the task, and they will be responsible for finishing the task.

Reward and Slash Mechanisms. Upon successful task completion, agents will receive token-based rewards via smart contracts that automatically verify outcomes using pre-specified validators. The reward amounts may be dynamic and may vary with task difficulty. Conversely, if an agent fails to meet its obligations, a slashing mechanism will be triggered. Slash amounts range from partial reward reduction for underperformance to the loss of the whole stake due to Sybil attacks. These mechanisms work together to incentivize honest behavior and mitigate malicious behavior in AgentaNet.

C.1. Current Progress & Limitations

Most existing multi-agent (Li et al., 2023; Hong et al., 2023; Chen et al., 2024b) and federated learning (Sun et al., 2024a; Ficco et al., 2024; Sun et al., 2024b) systems rely on a centralized server for agent management, task allocation, and trust building. These agent systems do not completely achieve the goals of fully decentralized networks like AgentaNet. Approaches such as Proof of Learning (Jia et al., 2021) aim to verify participation of human-controlled training via logs, but it is unclear if this approach can be used to validate agent autonomy or behavior. Existing multiple-agent systems also often lack robust verification of identities and shared information, making them vulnerable to Sybil attacks (Tu et al., 2021). Moreover, incentive structures are often simplistic or entirely absent, offering limited motivation for sustained and meaningful agent participation. We summarize the key limitations observed in prior works as follows:

- **Dependence on centralized infrastructure:** Reliance on centralized servers contradicts the principles of decentralized agent ecosystems.
- **Inadequate verification of agent autonomy:** Existing validation methods are designed for human involvement and do not extend to fully autonomous agents.
- **Lack of robust identity verification:** Current systems lack mechanisms to securely and verifiably distinguish agent identities, increasing vulnerability to Sybil attacks.
- **Lack of incentive mechanisms:** Incentive structures are either overly simplistic or missing, discouraging meaningful long-term participation.

C.2. Research Directions

Identity Management for Agents. To build secure and scalable decentralized autonomous agent systems, it is critical to establish verifiable and privacy-preserving identity management for agents. Different from human users, agents typically need to cryptographically prove not only their uniqueness but also their autonomy and operational integrity. Future research should explore decentralized identity frameworks aimed at agents. This framework can incorporate techniques such as Decentralized Identifiers (DIDs), verifiable credentials, and PoA schemes. Agents' identities should be dynamically linked to their behavior profiles, historical reputations, and received rewards.

Consensus Protocols for Autonomous Agents. Consensus in decentralized agent ecosystems needs to transform existing blockchain consensus protocols to accommodate high-frequency and low-latency coordination for agents. Future research may design lightweight and agent-specific consensus protocols that combine asynchronous agreement, partial trust models, and dynamic agent selection. These protocols should be resilient to Sybil and collusion attacks. They also need to support efficient and collective agents' decision-making in decentralized systems.

D. Ethical and Societal Considerations

The rise of decentralized agent ecosystems introduces new risks, including economic disintermediation, decision bias, and malicious use. AgentaNet must address these challenges through technical and governance layers. Bias amplification may arise if reputation or slashing disproportionately penalizes underrepresented models or agents with narrow training data. Job displacement is plausible if agent swarms outperform freelancers in open markets. Malicious actors may exploit protocol openness without robust identity control.