Toward an OSI-Inspired Stack for Agentic AI: From Fragmented Frameworks to Open Standards

Francisco-Javier Rodrigo-Ginés
NLP & IR Group
Universidad Nacional de Educación a Distancia (UNED)
Madrid, Spain
frodrigo@invi.uned.es

September 11, 2025

Abstract

Agentic Artificial Intelligence (Agentic AI) is emerging as a paradigm where autonomous agents, often powered by Large Language Models (LLMs), collaborate with tools, data sources, and each other to accomplish complex objectives. Yet the promise of these systems faces a fundamental obstacle: interoperability. Current frameworks and protocols remain fragmented, forcing costly ad hoc integrations and preventing the emergence of scalable, multi-agent ecosystems.

In this paper we argue that interoperability is not a secondary technical concern but the central challenge that will determine whether Agentic AI matures into robust digital infrastructure. To analyze this problem, we introduce an interoperability stack inspired by the OSI model in computer networking. The layered perspective clarifies responsibilities across seven levels, from physical infrastructure and runtime execution to agent-to-agent communication, coordination workflows, and governance and trust. Mapping today's initiatives onto this framework highlights stark asymmetries: lower layers are relatively mature, middle layers remain underdeveloped, and upper layers —especially governance— are almost entirely absent.

Building on this diagnosis, we outline a research and standardization agenda. Priorities include formal specifications for agent-to-agent communication, interoperable workflow schemas, benchmarks for measuring interoperability, and governance-by-protocol mechanisms that embed accountability and compliance into communication itself. We argue that, as with TCP/IP and the Internet, layered and open standards are the only path to prevent fragmentation and unlock the transformative potential of Agentic AI. Without them, the field risks repeating the stagnation of pre-Internet networking; with them, it may enable ecosystems of autonomous agents as consequential as the Internet itself.

***K*eywords** Agentic AI · Interoperability · Open Standards · OSI Model · Multi-Agent Systems · AI Governance

1 Introduction

Agentic Artificial Intelligence (Agentic AI) has recently emerged as one of the most promising paradigms in the field of Artificial Intelligence. Unlike conventional applications of Large Language Modelss (LLMs), which typically operate as isolated predictors of text, Agentic AI frames LLMs as autonomous agents capable of perceiving, planning, acting, and interacting in dynamic environments. These agents do not simply generate outputs; they can connect to external tools, query databases, call APIs, collaborate with other agents, and adapt their strategies in order to achieve complex objectives [1, 2]. In principle, such capabilities open the door to new forms of automation, decision support, and human–AI collaboration at scale.

Despite the rapid progress in demonstrations and prototypes, the field is at a critical juncture. The main bottleneck for Agentic AI is not model accuracy or reasoning ability per se, but rather integration and interoperability. Every potential deployment of agents—whether in enterprises, public administrations, or research institutions—must coexist with a heterogeneous ecosystem of pre-existing systems. These range from modern cloud infrastructures and REST APIs to deeply entrenched legacy systems built decades ago. In practice, organizations will not discard these infrastructures; instead, Agentic AI must adapt and integrate into them. Achieving this is a non-trivial challenge that will consume the majority of development and engineering effort in the coming years. Without robust interoperability, agent deployments will remain limited to isolated pilots or vendor-specific ecosystems, unable to scale into robust, reusable, and trustworthy infrastructures.

The history of computing shows that this challenge is not unique. Before the advent of standardized networking protocols, computer networks were fragmented and vendor-specific. Interoperability was extremely limited, and connecting heterogeneous systems required expensive, custom-built solutions. The turning point came with the introduction of the TCP/IP protocol suite in the 1970s [3] and the Open Systems Interconnection (OSI) reference model in the 1980s [4]. TCP/IP provided a practical, minimal set of interoperable protocols that enabled heterogeneous networks to communicate reliably. The OSI model, while not as widely deployed, offered a powerful conceptual framework that clarified responsibilities across seven layers, from physical transmission to application semantics. Together, these advances paved the way for the Internet: a modular, interoperable, and scalable ecosystem that became a global infrastructure.

We argue that Agentic AI today finds itself in a similar position to computer networking before TCP/IP and OSI. Multiple frameworks and initiatives exist—such as LangChain, AutoGen, Semantic Kernel, or vendor-specific orchestration solutions—but they remain siloed. Emerging protocols like the Model Context Protocol (MCP) [5] or proposals for Agent-to-Agent (A2A) [6] communication are promising but partial. There is no unifying framework that clarifies what interoperability means across different layers of an agentic ecosystem, nor how to evaluate and enforce it. As a result, the current landscape is fragmented, innovation is hindered by incompatibilities, and organizations face high costs and risks of vendor lock-in.

This paper addresses these gaps by proposing an OSI-inspired interoperability stack for Agentic AI. Our approach is not to introduce yet another protocol, but rather to provide a layered conceptual model that organizes the problem space. By mapping interoperability challenges onto layers—from infrastructure and runtime execution at the bottom to coordination and governance at the top—we aim to achieve three goals:

1. Clarify the responsibilities of each layer and the distinct forms of interoperability required.
2. Situate existing initiatives (e.g., MCP, A2A, orchestration frameworks) within this layered model, highlighting what they solve and what they leave unaddressed.
3. Identify research gaps and propose an agenda for developing open, layered standards that can transform fragmented prototypes into robust agentic ecosystems.

The contribution of this paper is thus twofold. First, it introduces an OSI-inspired conceptual framework that makes explicit the layered nature of interoperability in Agentic AI. Second, it outlines a research and standardization agenda to close the most pressing gaps, focusing on three dimensions: (i) formal specifications for agent-to-agent communication, (ii) benchmarks and metrics for interoperability, and (iii) governance-by-protocol mechanisms for accountability and compliance.

The rest of this paper is organized as follows. Section 2 reviews the background, drawing connections between Multi-Agent Systems (MAS) research, classical agent communication protocols, and the new dynamics introduced by LLM-based agents. Section 3 defines interoperability as the central challenge for Agentic AI, distinguishing its technical, semantic, and governance dimensions. Section 4 presents the OSI-inspired interoperability stack, situating current initiatives across its seven layers. Section 5 discusses the research gaps this framework reveals and proposes an agenda for closing them. Finally, Section 6 concludes with implications for future research, standardization, and practice.

2 Background and related work

The vision of autonomous agents operating in open and heterogeneous environments is not new. Long before the advent of Large Language Models, the field of Multi-Agent Systems (MAS) investigated how collections of interacting agents could cooperate, negotiate, and compete in order to solve problems that exceeded the capabilities of any individual system [7, 8]. MAS became a central strand of distributed Artificial Intelligence, with applications ranging from robotics and logistics to simulation and electronic commerce.

2.1 Classical MAS and communication protocols

In MAS research, one of the core challenges was precisely interoperability. Agents were often developed by different teams, in different languages, and with different knowledge representations. To enable collaboration, standard communication protocols were proposed.

The Knowledge Query and Manipulation Language (KQML) [9] was an early attempt to create a message-based communication standard. It defined a set of performatives—such as `ask`, `tell`, and `achieve`—inspired by speech act theory. KQML allowed agents to exchange information about beliefs, goals, and intentions, but it struggled with semantic mismatches and scalability.

The Foundation for Intelligent Physical Agents (FIPA) took these ideas further with the Agent Communication Language (FIPA-ACL) [10]. FIPA-ACL specified message structures, performatives, and semantics grounded in formal models of mental states. Combined with interaction protocols (e.g., for auctions, negotiations, or task allocation), FIPA-ACL became the de facto standard in academic MAS research. Other initiatives, such as the Agent Communication Protocol (ACP) [11], explored lightweight or domain-specific alternatives.

Despite their elegance, these standards remained mostly confined to research settings. There were three main limitations. First, they relied heavily on symbolic reasoning and predefined ontologies, which made them brittle in practice. Second, semantic mismatches between agents were difficult to resolve automatically. Third, adoption in industry was limited: implementing full FIPA-compliant infrastructures was costly, and integration with legacy enterprise systems remained cumbersome. As a result, MAS never achieved widespread deployment outside specialized domains.

2.2 The shift introduced by LLMs

The rise of LLMs has radically changed the picture. Unlike symbolic agents, LLM-based agents can use natural language as both an interface and a medium of reasoning. They can parse unstructured data, generate flexible outputs, and interact with humans and machines through natural language instructions. This shift makes it possible to instantiate multi-agent simulations and workflows with minimal engineering effort, as shown by recent demonstrations of generative agents [1] and reasoning–acting frameworks such as ReAct [2].

What has emerged is often referred to as Agentic AI [12, 13, 14]. Agentic AI combines the autonomy and coordination principles of classical MAS with the generative, adaptive, and general-purpose capabilities of foundation models. This hybrid paradigm enables new forms of interaction:

- **Probabilistic reasoning**: agents reason not through strict logical inference, but via probabilistic, generative mechanisms.
- **Emergent behavior**: coordination patterns can emerge from interactions without being explicitly programmed.
- **Rapid instantiation**: agents can be composed into multi-agent systems on-the-fly using high-level natural language instructions.

Yet these advances also revive old challenges. Like their MAS predecessors, LLM-based agents must operate in heterogeneous environments. They must integrate with APIs, databases, enterprise applications, and increasingly with each other. Without interoperability, each agentic framework remains a silo.

2.3 Contemporary initiatives and fragmentation

Several initiatives have emerged to tackle pieces of the interoperability problem:

- **Model Context Protocol (MCP)**: proposed by Anthropic, MCP standardizes how models discover and invoke external tools, somewhat analogous to how OpenAPI structured web services.
- **Agent-to-Agent (A2A)** protocols: still in early stages, these proposals aim to enable discovery, negotiation, and cooperation among agents, recalling the ambitions of FIPA-ACL but adapted to LLM-driven interactions.
- **Frameworks like LangChain, AutoGen, CrewAI, Semantic Kernel**: these provide orchestration, memory, and workflow management, but do not define interoperable standards across implementations. An agent built in one framework cannot easily collaborate with an agent in another.

While promising, these efforts remain fragmented. There is no comprehensive architecture that integrates tool interoperability, agent-to-agent communication, and governance requirements. Each initiative solves a local problem but leaves systemic interoperability unaddressed.

2.4 Learning from networking history

This situation resembles the state of computer networking in the 1970s. Networks existed, but they were vendor-specific and could not interoperate. Each connection required bespoke engineering. The breakthrough came with the development of TCP/IP [3], which offered a minimal, practical set of interoperable protocols, and the OSI reference model [4], which provided a conceptual framework to structure communication into seven layers.

Although OSI was not as widely deployed as TCP/IP, it became a powerful analytical tool. It clarified responsibilities: physical transmission, data link, network routing, transport reliability, session management, presentation semantics, and application logic. Crucially, it showed that interoperability is not a monolithic problem but a layered one.

We argue that Agentic AI today occupies a position comparable to the early days of computer networking. Multiple frameworks and protocols already exist, and in some cases they can interoperate or be bridged, but such connections remain partial and ad hoc. What the field still lacks is a layered reference model that clarifies responsibilities, identifies where existing initiatives fit, and highlights the gaps that remain unaddressed. By adopting an OSI-inspired perspective, interoperability can be structured in a systematic way, preventing duplication of fragmented solutions and guiding the evolution of open, layered standards.

3 Interoperability as the central challenge

Interoperability can be broadly defined as the capacity of heterogeneous systems to exchange information, interpret it meaningfully, and act upon it coherently. In the context of Agentic AI, interoperability is not a peripheral concern but the central bottleneck that will determine whether agent ecosystems remain isolated prototypes or evolve into robust infrastructures comparable to the Internet.

3.1 Dimensions of interoperability

We distinguish three complementary dimensions of interoperability in Agentic AI:

Tool interoperability. Agents must be able to access and use external resources—APIs, databases, enterprise applications—in a consistent manner. At present, every integration is a bespoke engineering effort. Frameworks like LangChain have simplified tool calling within a controlled environment, but there is no universal standard comparable to OpenAPI for services. Without tool interoperability, organizations face duplicated work and increased fragility when switching vendors or upgrading infrastructures.

Agent-to-agent interoperability. Autonomous agents should be able to discover each other, negotiate roles, and collaborate across heterogeneous frameworks and vendors. Today, an agent developed in LangChain cannot directly communicate with one in AutoGen or CrewAI, unless a human developer manually engineers a bridge. This is analogous to the pre-TCP/IP era of networking, when networks existed but were siloed. Without protocols for discovery and communication, the vision of agent ecosystems remains unrealized.

Governance interoperability. As agents begin to operate in sensitive domains such as healthcare, finance, or public administration, interoperability must extend beyond technical communication into governance. Questions of permissioning, auditing, liability, and regulatory compliance (e.g., GDPR or the EU AI Act) cannot be solved by individual frameworks. Instead, they require shared governance mechanisms embedded into protocols themselves. Without such mechanisms, organizations will lack the trust to deploy Agentic AI at scale.

3.2 Risks of fragmentation

The absence of interoperability generates several systemic risks:

- **Increased integration costs.** Each deployment requires custom engineering to connect frameworks, tools, and legacy systems. This slows innovation and creates barriers to entry.
- **Vendor lock-in.** Proprietary frameworks dominate, forcing organizations to commit to one ecosystem. Switching costs rise, innovation decreases, and monopolistic dynamics emerge.
- **Limited trust and accountability.** Without standardized governance mechanisms, organizations cannot ensure that agents behave in compliant and auditable ways. This is especially problematic in regulated industries.
- **Inhibited innovation.** Just as the absence of networking standards limited the growth of early computer networks, the lack of interoperability will prevent Agentic AI from scaling beyond isolated prototypes.

3.3 The case for layered interoperability

The analogy with networking highlights why interoperability should be viewed as layered rather than monolithic. In the 1970s, different networking technologies existed—ARPANET, X.25, vendor-specific LANs—but they could not interoperate. Each integration required custom solutions. The introduction of TCP/IP provided a minimal set of interoperable protocols that scaled globally, while the OSI model clarified responsibilities across seven layers. This layered approach enabled modularity: progress at one layer (e.g., physical transmission) could propagate upward without disrupting higher layers (e.g., applications).

Agentic AI today faces a similar situation. Lower layers, such as infrastructure (cloud, GPUs) and runtime execution (API calls, containers), are relatively mature. But higher layers, especially agent-to-agent protocols, coordination semantics, and governance, remain underdeveloped. Without a layered framework, discussions of interoperability remain fragmented: MCP addresses model-to-tool invocation, while A2A initiatives propose communication protocols, but neither situates itself within a larger architecture.

3.4 Why interoperability is the core development effort

We argue that interoperability will consume the majority of engineering and research effort in Agentic AI. Developing a new LLM or fine-tuning a model is valuable, but organizations already have access to a growing ecosystem of strong models. The real challenge is integrating these agents into complex, heterogeneous environments where existing infrastructures cannot simply be replaced.

For enterprises, this means ensuring that AI agents can access ERP systems, customer databases, and analytics platforms without bespoke integration each time. For public administrations, it means agents must interact with legacy registries, citizen portals, and cross-agency systems. In both contexts, interoperability is the key determinant of whether Agentic AI provides value or remains confined to pilot projects.

Thus, interoperability is not merely a technical requirement but the cornerstone of sustainable Agentic AI. It is where the most significant barriers lie, where the most resources will be invested, and where the success or failure of agent ecosystems will ultimately be determined.

4 An OSI-inspired interoperability stack

The layered approach inspired by the OSI reference model provides a structured way to analyze interoperability in Agentic AI. Instead of treating integration as a monolithic issue, it allows us to disentangle the different responsibilities, situate current initiatives, and identify where the most critical gaps remain. The proposed mapping is summarized in Table 1, which provides an overview of the seven layers, their equivalents in the Agentic AI context, and illustrative examples. In the remainder of this section, we expand on each layer in detail, discussing its current state, limitations, and future directions for research and industry.

4.1 Layer 1: Physical (Infrastructure)

At the foundation of the stack lies the physical layer, which encompasses the computational infrastructure and connectivity on which agentic systems operate. This includes GPUs and accelerators, distributed storage, high-speed networking, and cloud services that host LLMs and orchestrators. In one sense, this layer appears relatively mature: hyperscale cloud providers have built reliable infrastructures for training and serving large models, and containerization technologies offer portability across environments.

Yet the picture is more complex when viewed through the lens of interoperability. Hybrid deployments that combine cloud services with on-premise infrastructures are common, but they depend on proprietary APIs and bespoke bridges. Migrating workloads between providers is costly and fragile, reinforcing lock-in. Moreover, the concentration of GPU capacity in a handful of companies creates structural dependencies, raising geopolitical and economic risks.

Another overlooked dimension is sustainability. The physical layer consumes enormous energy, and interoperability discussions rarely integrate metrics for carbon footprint or green optimization. In the future, interoperability standards could embed sustainability indicators, enabling organizations to choose infrastructures not only for cost and performance but also for ecological impact. This would mirror how energy efficiency standards became integral to networking hardware.

Table 1: OSI-inspired interoperability stack for Agentic AI, with concrete responsibilities, current examples, and future directions.

OSI layer	Agentic AI equivalent	Responsibilities and examples	Gaps and future directions
Application	Governance & Trust	Embedding accountability, permissions, compliance, and auditability. Currently handled ad hoc through organizational policies; no standard protocol support.	Develop "governance-by-protocol" mechanisms (e.g., mandatory audit fields in agent messages, GDPR-aware tokens, permission identifiers). Explore blockchain for immutable logs and ZK proofs for compliance checks.
Presentation	Coordination & Workflows	Defining workflows, delegating subtasks, and aligning multiple agents toward collective goals. Current examples: LangGraph for workflow orchestration, CrewAI for team-based delegation, Semantic Kernel for planning.	Lack of portable workflow schemas; no shared semantics for consensus or conflict resolution. Future work: formal workflow standards akin to BPMN, enabling workflows to be reused across frameworks.
Session	Agent-to-Agent Protocols	Enabling discovery, negotiation, dialogue management, and cooperation between agents. Early A2A proposals exist; MAS precedents include FIPA-ACL and KQML.	No widely adopted session protocol; discovery and negotiation remain ad hoc. Future direction: hybrid protocols combining structured performatives (request, inform, confirm) with natural-language grounding.
Transport	Reliable Messaging	Guaranteeing delivery, ordering, and integrity of agent communications. Current examples: gRPC, Kafka, message queues used in orchestration.	Lack of guarantees tailored to long conversations, streaming outputs, and probabilistic responses. Future work: secure transport protocols adapted to LLM agents, embedding authentication and encryption by default.
Network	Model-to-Tool Protocols	Standardizing discovery and invocation of tools and APIs. Example: Model Context Protocol (MCP). Analogy: OpenAPI for web services.	Fragmented adoption; inconsistent tool descriptions and metadata. Future work: universal schemas for tool capabilities, standardized versioning and permissioning, neutral consortium to prevent fragmentation.
Data Link	Runtime Execution	Handling execution environments and function calls. Examples: OpenAI function calling, LangChain tool wrappers, containerized microservices.	Portability issues across frameworks; reproducibility and auditing absent. Future work: portable runtime specifications analogous to POSIX; embedded auditing of function calls to enable accountability.
Physical	Infrastructure	Underlying compute and connectivity: GPUs, accelerators, distributed storage, hybrid cloud/on-premise systems.	Strong concentration of GPU capacity in a few vendors; limited portability across infrastructures; sustainability not considered. Future work: interoperability for hybrid infrastructures and integration of sustainability metrics into deployment standards.

4.2 Layer 2: Data Link (Runtime Execution)

The data link layer addresses how agents execute actions in runtime environments. It covers the mechanics of invoking APIs, binding functions, and encapsulating services within containers or microservices. In practice, this is where an LLM agent translates an intention into an executable call.

At present, frameworks offer partial solutions. Function calling in OpenAI's API, LangChain's tool wrappers, and AutoGen's delegation mechanisms all provide execution pathways. However, they lack portability: each defines its own format for functions, argument schemas, and error handling. This creates brittle integrations and hampers reuse.

Another gap lies in reproducibility and auditability. Two executions of the same function by different agents may produce different outcomes if runtime environments differ subtly. Without standardization, tracing these discrepancies becomes impossible. A lesson can be drawn from POSIX [15] in operating systems: a minimal set of standardized calls provided consistency across platforms, enabling portability and debugging. Agentic AI requires a similar effort for runtime execution.

Future work should explore portable runtime specifications, common invocation schemas, and embedded auditing mechanisms. This would not only improve reliability but also enable accountability: organizations could trace how, when, and why a function was invoked, ensuring compliance with internal and external requirements.

4.3 Layer 3: Network (Model-to-Tool Protocols)

The network layer manages how agents discover and use external tools. This is the realm of protocols that allow an LLM agent to know what tools are available, what their capabilities are, and how to invoke them.

The most significant initiative here is the Model Context Protocol (MCP), proposed by Anthropic. MCP specifies how tools can advertise their capabilities and how models can invoke them, somewhat akin to how OpenAPI structured web services. While promising, MCP is not yet universal. Adoption is limited, and frameworks implement tool invocation in different ways.

A key gap is the lack of common metadata schemas. Tools may be described inconsistently, making automatic discovery unreliable. Versioning, error handling, and permission management are also largely absent. Without these elements, interoperability is brittle: an agent that learns to use a tool via MCP in one ecosystem may not transfer that ability to another.

Moving forward, research should formalize schemas for tool discovery and description, inspired by precedents such as GraphQL [16] or schema.org [17]. Industry consortia could play a role in converging standards, preventing the proliferation of incompatible variants. Importantly, permissioning and trust should be embedded at this layer, ensuring that tools cannot be invoked without proper authorization.

4.4 Layer 4: Transport (Reliable Messaging)

Transport protocols ensure that communication between agents, orchestrators, and tools is reliable. In classical networking, this role was played by TCP, which guaranteed delivery, ordering, and error correction, providing the foundation on which higher-level interoperability could flourish. In Agentic AI, transport protocols must fulfill a similar role: ensuring that messages between agents are delivered despite probabilistic outputs, long-running contexts, and dynamic environments. The challenge is not merely one of message passing, but of sustaining coherence across conversations that unfold incrementally, involve reasoning steps, and may include streaming outputs that extend over time.

At present, most frameworks rely on generic middleware such as message queues, gRPC, or Kafka. These technologies offer useful primitives but were never designed for agentic scenarios. As a result, they introduce critical limitations: they lack native support for long conversational contexts, do not address the probabilistic nature of LLM-generated messages, and often struggle with incremental or streaming communication patterns. Error handling is also fragmented, depending on ad hoc strategies such as retries or heuristics, with little standardization across systems. This fragility undermines interoperability, as each framework must reinvent basic mechanisms of reliability.

Security represents another major gap. Current approaches to messaging in agentic systems rarely include built-in mechanisms to authenticate agents, prevent tampering, or guarantee confidentiality. In practice, this leaves multi-agent communication vulnerable to interception, replay, or manipulation by malicious actors. The lack of standardized authentication and encryption protocols not only hampers trust between heterogeneous agents but also limits the safe deployment of agentic systems in sensitive or regulated environments.

Emerging initiatives illustrate that the community is beginning to recognize this gap. Protocols such as the Model Context Protocol (MCP), Agent Communication Protocol (ACP), Agent-to-Agent (A2A) communication, and Agent Networking Protocol (ANP) represent early steps toward formalizing reliable message exchange. MCP and ACP adopt a more centralized or brokered architecture, focusing on message structure and context management. A2A enables direct peer-to-peer interactions, while ANP, developed by Cisco, explores decentralized topologies for communication and coordination across large networks of agents. These efforts remain fragmented, but together they highlight an increasing awareness that transport protocols must evolve to handle discovery, routing, reliability, and security in ways tailored to agentic communication.

By analogy with the Internet, one might argue that Agentic AI today lacks the equivalent of TCP/IP. Just as TCP provided the reliability guarantees that enabled interoperable higher-layer protocols such as HTTP or SMTP, future transport protocols for agents will need to incorporate state synchronization, retries, streaming guarantees, and secure authentication. Such protocols must be lightweight enough to operate across diverse platforms, yet robust enough to handle dynamic environments and adversarial conditions. Their adoption would establish the foundation for higher-layer interoperability, allowing currently siloed frameworks—such as LangChain, AutoGen, and CrewAI—to communicate without bespoke bridges or ad hoc middleware. In this sense, reliable transport is not merely a technical detail but a structural prerequisite for an open and robust Agentic AI ecosystem.

4.5 Layer 5: Session (Agent-to-Agent Protocols)

The session layer manages sustained interactions between agents. It is here that agents should be able to discover one another, initiate dialogues, negotiate roles, and establish cooperative or competitive relationships.

At present, this is the most underdeveloped layer. There are preliminary proposals for agent-to-agent (A2A) protocols, but no consensus or adoption. Discovery is manual, negotiation absent, and cooperation pre-programmed. Without session protocols, heterogeneous agents cannot meaningfully collaborate.

This recalls the ambitions of MAS research in the 1990s, when FIPA-ACL and KQML attempted to define performatives for agent communication. Those efforts faltered in practice because they were too rigid, requiring agents to share symbolic ontologies. LLM-based agents introduce new possibilities: negotiation and cooperation could occur in natural language, while structured elements provide grounding.

Research must therefore explore hybrid approaches. Structured performatives could define basic actions (request, inform, confirm), while natural language fills in semantics. Industry must align on interoperable A2A protocols, ideally through neutral organizations. Without such protocols, multi-agent ecosystems will remain fragmented.

4.6 Layer 6: Presentation (Coordination and Workflows)

Coordination and workflow orchestration occur at the presentation layer. Here, the concern is not just message passing, but aligning multiple agents toward collective goals.

Frameworks like LangGraph, CrewAI, or Semantic Kernel provide orchestration capabilities: memory management, task decomposition, delegation. These are powerful but vendor-specific. Workflows defined in one cannot be ported to another. Coordination semantics, such as consensus, conflict resolution, or task prioritization, are often implicit or absent.

The gap is striking: without shared schemas for workflows, organizations cannot reuse or transfer agentic processes. This mirrors the pre-standardization era of business process modeling, before BPMN [18] provided a portable notation. A similar effort is required here: formal workflow schemas for agentic coordination, enabling portability and reuse.

Research should formalize coordination strategies, while industry defines workflow standards. The result would be an ecosystem where workflows could circulate across organizations, much as BPMN processes do today. This would reduce lock-in and accelerate adoption.

4.7 Layer 7: Application (Governance and Trust)

At the top of the stack lies governance and trust. This is not about technical integration but about embedding accountability, compliance, and trustworthiness into the fabric of agentic interaction.

Today, this layer is almost entirely absent. Governance is treated as an organizational add-on rather than a protocol-level concern. Yet without it, adoption in sensitive domains is unlikely. In healthcare, finance, or government, organizations require guarantees: who authorized an action, what permissions were checked, how compliance was enforced. Without embedded governance, trust collapses.

The gap is profound: no standard exists for auditability, liability, or compliance in agentic protocols. This must change. Research should develop governance-by-protocol mechanisms that embed traceability, permissioning, and accountability into every interaction. Regulators and industry must collaborate to ensure alignment with legal frameworks such as the EU AI Act or GDPR.

Embedding governance in protocols will not only ensure compliance but also enable interoperability of trust. Just as TLS made secure communication standard in networking, governance mechanisms could make trustworthy agentic interaction the norm rather than the exception.

5 Discussion

The OSI-inspired stack provides a useful map of the interoperability problem in Agentic AI. By decomposing the challenge into layers, it becomes clear that the difficulties facing the field are not uniform. Some layers are relatively mature, supported by decades of advances in computing and distributed systems, while others are almost entirely absent from current discourse. This asymmetry is critical because it helps us prioritize where research, development, and standardization should focus in the years ahead.

5.1 Asymmetries across layers

The lower layers of the stack, corresponding to infrastructure, runtime execution, and model-to-tool protocols, are comparatively well developed. Cloud computing, containerization, and middleware already provide robust building blocks. Efforts like the Model Context Protocol show that industry is beginning to converge on standards for tool invocation, even if adoption remains partial. In contrast, the middle layers—reliable messaging, agent-to-agent protocols, and coordination workflows—are fragmented and immature. These layers are crucial because they define how agents interact with each other, sustain dialogues, and align toward shared goals. Without them, Agentic AI remains confined to isolated pilots or single-vendor ecosystems. At the top, the application layer of governance and trust is almost completely absent, despite being essential for adoption in regulated domains.

This asymmetry suggests a paradox. The field is progressing quickly in areas that are easiest to tackle—deploying models in clouds, invoking APIs—but it is neglecting the layers that matter most for large-scale, sustainable adoption. The danger is that innovation will stall as organizations encounter the limits of bespoke integration, just as computer networking once stagnated before the adoption of TCP/IP.

5.2 Research gaps

Several gaps stand out as priorities for the research community:

First, there is no shared vocabulary or formalism for evaluating interoperability. In the absence of metrics, claims of interoperability are anecdotal, based on isolated demonstrations rather than systematic comparison. Benchmarks for interoperability, akin to GLUE [19] or MMLU [20] for LLM reasoning, would enable progress to be measured and compared. For example, tasks could be designed to test whether two heterogeneous agents can discover each other, negotiate a protocol, and complete a joint task without bespoke engineering.

Second, security is almost entirely missing from current discussions. Reliable transport in networking only succeeded when protocols incorporated authentication, integrity, and encryption. Similarly, multi-agent systems will require defenses against adversarial behaviors, such as message injection, impersonation, or denial-of-service attacks at the agentic layer. Research must adapt security primitives to the probabilistic and natural-language-driven nature of LLMs, ensuring that interoperability does not open new attack vectors.

Third, governance requires conceptual and technical innovation. Embedding compliance and accountability into protocols will not be trivial. It demands new models of "governance-by-protocol", where permissions, audit trails, and liability are encoded in the very messages agents exchange. This raises open questions: how should accountability be distributed among agents? How can compliance be verified automatically? How can privacy rights be preserved when agents share context? These are pressing research challenges that must be addressed in collaboration with legal scholars and regulators.

5.3 Industry challenges

The challenges for industry are complementary. Vendors today are locked in a competitive race to capture market share, leading to proliferation of incompatible frameworks. While innovation is valuable, unchecked divergence risks repeating the mistakes of pre-standardization networking, where proprietary protocols hindered interoperability.

There is also the challenge of incentives. Open standards often run against the immediate interests of dominant vendors, who benefit from customer lock-in. Overcoming this requires coordinated pressure from regulators, customers, and consortia. For example, the role played by IETF in the Internet or W3C in the Web shows that neutral bodies can steer industry toward open protocols without stifling innovation. A similar initiative—an “IETF for Agentic AI”—may be required to shepherd the development of A2A standards, workflow schemas, and governance mechanisms.

Another industry challenge is sustainability. As discussed in the physical layer, the carbon footprint of large-scale deployments is immense. Industry adoption of interoperability standards could incorporate sustainability metrics, enabling organizations to make choices that are not only technically sound but also ecologically responsible.

5.4 Toward a research and standardization agenda

Bringing these observations together, we propose an agenda with three main axes.

The first is technical formalization. Research should produce specifications for agent-to-agent protocols, interoperable workflow schemas, and governance-by-protocol mechanisms. These should be designed with modularity in mind, allowing incremental adoption and adaptation to different domains. Lessons from MAS research—such as the performative structure of FIPA-ACL—can provide inspiration, but they must be adapted to the probabilistic and generative nature of LLM-based agents.

The second axis is evaluation. The field urgently needs benchmarks and metrics for interoperability. Shared evaluation tasks would not only measure progress but also create incentives for convergence: vendors will seek to demonstrate compliance with benchmarks, much as they do with reasoning or robustness tests today. Establishing such benchmarks could be the role of academic consortia or independent initiatives, ideally involving both researchers and practitioners.

The third axis is institutional. Open standards require neutral bodies to maintain them, foster adoption, and balance interests. Without such institutions, the field risks fragmentation. Whether through existing organizations (ISO/IEC, IEEE, W3C) or new ones, the establishment of a dedicated forum for Agentic AI interoperability is indispensable. This forum should include not only vendors and researchers, but also regulators, civil society, and domain experts from sensitive sectors such as healthcare and finance.

5.5 The centrality of interoperability

The overarching message is that interoperability is not just another technical detail, but the central challenge for Agentic AI. It is where most of the engineering hours will be spent, where the most significant risks lie, and where the future of the field will be decided. Model development and fine-tuning are important, but they are not sufficient. The real bottleneck is whether agents built in different frameworks, by different organizations, and for different purposes can meaningfully communicate, coordinate, and comply with shared governance.

The history of networking teaches us that layered, open standards can unlock global transformation. Without them, Agentic AI will remain a patchwork of silos; with them, it can become as transformative as the Internet. The task for research and industry is therefore not merely to build better agents, but to build better ways for agents to interoperate.

6 Conclusion and future work

This paper has argued that interoperability is not a peripheral concern but the central challenge of Agentic AI. The enthusiasm surrounding autonomous agents powered by large language models often focuses on their capacity for reasoning, planning, or emergent behaviors. Yet these capabilities will remain largely academic curiosities or isolated prototypes unless agents can be integrated into the complex, heterogeneous ecosystems that characterize real-world organizations. Enterprises, public administrations, and research institutions all operate on infrastructures that cannot simply be replaced: they must be bridged. Interoperability is the condition of possibility for this bridging, and thus for the very viability of Agentic AI at scale.

To analyze this problem, we introduced an OSI-inspired interoperability stack. By decomposing interoperability into seven layers—from physical infrastructure to governance and trust—the model clarifies the different responsibilities that must be addressed. It also situates current initiatives within a broader architecture. Lower layers such as infrastructure and runtime execution are relatively mature, while tool invocation protocols like MCP show promising convergence at the network layer. But the middle layers, encompassing transport, agent-to-agent communication, and coordination workflows, remain fragmented and underdeveloped. At the top, governance and trust are almost absent, despite being essential for adoption in sensitive domains. This layered perspective provides not only a conceptual map but also a diagnostic tool: it reveals where progress has been made, where gaps remain, and where research and industry must invest their efforts.

The implications of this perspective extend across academia, industry, and regulation. For researchers, the stack highlights an agenda of open problems: the need for formal specifications of agent-to-agent protocols, for interoperable workflow schemas, and for governance-by-protocol mechanisms that embed accountability into communication itself. It also points to the urgency of benchmarks: without systematic ways of evaluating interoperability, progress will remain anecdotal. For industry, the stack underscores the danger of fragmentation. Vendors face strong incentives to lock customers into proprietary ecosystems, but such short-term gains risk reproducing the stagnation of pre-Internet networking. Open standards are the only way to unlock modular ecosystems that scale globally. Finally, for regulators and policymakers, the stack shows that governance cannot be an afterthought. Just as security was embedded into transport protocols, compliance and accountability must be embedded into agentic protocols if they are to be trustworthy in domains such as healthcare or finance.

Looking ahead, several directions for future work emerge. First, the field must pursue technical formalization. This includes adapting insights from MAS protocols such as FIPA-ACL and KQML to the probabilistic and natural-language-driven context of LLM-based agents, while avoiding the rigidity that limited their adoption. It also includes defining schemas for workflow coordination that allow processes to be portable across frameworks, much as BPMN did for business processes. Second, the field must establish evaluation frameworks. Benchmarks for interoperability could test whether agents built in different ecosystems can discover each other, exchange structured and unstructured messages, and complete tasks collaboratively. Such benchmarks would not only measure progress but also create incentives for convergence, as vendors would be motivated to demonstrate compliance. Third, governance must move from organizational policy to protocol-level mechanisms. This entails embedding permissions, audit trails, and compliance checks directly into communication standards, ensuring that trust is not an external layer but an intrinsic property of agentic interaction.

Future work also demands institutional innovation. History shows that open standards do not emerge spontaneously: they require neutral bodies to maintain them, foster adoption, and balance competing interests. Just as the Internet benefited from the IETF and the Web from the W3C, Agentic AI will likely require a dedicated forum for interoperability standards. Such a forum should not be limited to vendors and researchers but should include regulators, civil society, and domain experts from sectors where trust is paramount. The creation of such institutions may be as important as any technical advance, for without them, incentives for openness will remain weak.

The lesson from networking history is clear. Before TCP/IP, networks existed but were fragmented, costly, and incompatible. It was only when layered, open protocols were adopted that the Internet emerged as a global infrastructure. Agentic AI stands at a similar crossroads. Without standards, it risks remaining a patchwork of incompatible frameworks and bespoke integrations. With them, it could enable ecosystems of autonomous agents that are as transformative as the Internet itself. The challenge is immense, but so too is the opportunity. If interoperability can be achieved—not as an afterthought but as the central axis of development—Agentic AI may fulfill its promise as a new layer of digital infrastructure, reshaping how humans, organizations, and machines interact.

References

[1] Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pages 1–22. ACM, 2023.

[2] Shunyu Yao, Jeffrey Zhao, Dian Yu, Karthik Narasimhan, Jeanne Cao, et al. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2023.

[3] Vinton G. Cerf and Robert E. Kahn. A protocol for packet network intercommunication. In *IEEE Transactions on Communications*, volume 22, pages 637–648. IEEE, 1974.

[4] Hubert Zimmermann. Osi reference model–the iso model of architecture for open systems interconnection. *IEEE Transactions on Communications*, 28(4):425–432, 1980.

[5] Xinyi Hou, Yanjie Zhao, Shenao Wang, and Haoyu Wang. Model context protocol (mcp): Landscape, security threats, and future research directions. *arXiv preprint arXiv:2503.23278*, 2025.

[6] Partha Pratim Ray. A review on agent-to-agent protocol: Concept, state-of-the-art, challenges and future directions. *Authorea Preprints*, 2025.

[7] Michael Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2nd edition, 2009.

[8] Nicholas R. Jennings, Katia Sycara, and Michael Wooldridge. Agent-based software engineering. *International Journal of Automated Software Engineering*, 5(1):5–46, 1998.

[9] Tim Finin, Yannis Labrou, and James Mayfield. Kqml as an agent communication language. In *Proceedings of the 3rd International Conference on Information and Knowledge Management (CIKM)*, pages 456–463. ACM, 1994.

[10] Foundation for Intelligent Physical Agents. Fipa acl message structure specification. Technical report, FIPA, 2002. Available at `http://www.fipa.org/specs/fipa00061/`.

[11] James Odell, H. Van Dyke Parunak, and Bernhard Bauer. Agent communication languages: Acl and beyond. In *Agent-Oriented Software Engineering II*, pages 73–92. Springer, 2002.

[12] Xueqing Wang, Hanze Zhang, Canwen Xu, Huan Xu, Hongyin Li, Wei Sun, et al. A survey on large language model-based autonomous agents. *arXiv preprint arXiv:2308.11432*, 2023.

[13] Zhen Xi, Mengnan Li, Wen Zhou, et al. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*, 2023.

[14] Meta AI Research. Agentic ai: Research frontiers and open problems. Technical report, Meta, 2024. White Paper, available at `https://ai.meta.com/research/`.

[15] Stephen R Walli. The posix family of standards. *StandardView*, 3(1):11–17, 1995.

[16] Antonio Quiña-Mera, Pablo Fernandez, José María García, and Antonio Ruiz-Cortés. Graphql: A systematic mapping study. *ACM computing surveys*, 55(10):1–35, 2023.

[17] Phil Barker and Lorna M Campbell. What is schema. org. *LRMI. Retrieved April*, 21:2015, 2014.

[18] Stephen A White. Introduction to bpmn. *Ibm Cooperation*, 2(0):0, 2004.

[19] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018.

[20] Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, et al. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems*, 37:95266–95290, 2024.