Upgrade or Switch: The Need for New Registry Architecture for the Internet of AI Agents

Ramesh Raskar, Pradyumna Chari, Jared Grogan, Mahesh Lambe, Dimitris Stripelis, Robert Lincourt, Rajesh Ranjan, Shailja Gupta, Surya Narayan Singh

Project NANDA

Introduction

The web is on the cusp of a profound transformation. While the current World Wide Web primarily connects humans to information and services via static and dynamic web pages, websites, and APIs that respond only when explicitly called upon, the emerging "Internet of AI Agents" promises to connect humans with autonomous digital agents capable of performing tasks, making decisions, and interacting with other agents on our behalf.

Unlike traditional web components that remain inert until activated by human requests, these AI agents are persistent, proactive computational entities with built-in reasoning capabilities that can anticipate needs, take initiative, maintain ongoing state, and work toward defined goals without constant human direction. They leverage advanced machine learning models to interpret ambiguous instructions, adapt to changing circumstances, and make context-sensitive decisions within their domain of operation - capabilities fundamentally different from the request-response pattern of conventional web architecture.

These agents - powered by advanced AI and operating with varying degrees of autonomy - will fundamentally change how we interact with digital systems [2] and with each other through digital intermediaries. Figure 1 shows an example.

The rise of autonomous agents presents a significant challenge to our current internet infrastructure. Originally designed for human-driven, browser-based interactions, systems like DNS, IP addressing, and identity verification were not built to support the scale, speed, and security demands of billions of automated agents communicating and transacting in real time. As of 2025, there are over 1.1 billion websites, yet only about 193 million (17%) are actively maintained . Meanwhile, the internet serves 5.56 billion users worldwide, and there are approximately 7.21 billion smartphones in use [1] . This vast and rapidly growing digital ecosystem underscores the need for a more robust and scalable infrastructure to support the emerging network of autonomous agents.

This paper examines whether we should upgrade the existing web infrastructure or switch to entirely new registry architectures specifically designed for the internet of AI agents, or agentic web. By examining historical technology transitions and analyzing the known and unknown challenges ahead, we aim to provide a roadmap for the infrastructure needed to realize the full potential of autonomous digital agents.

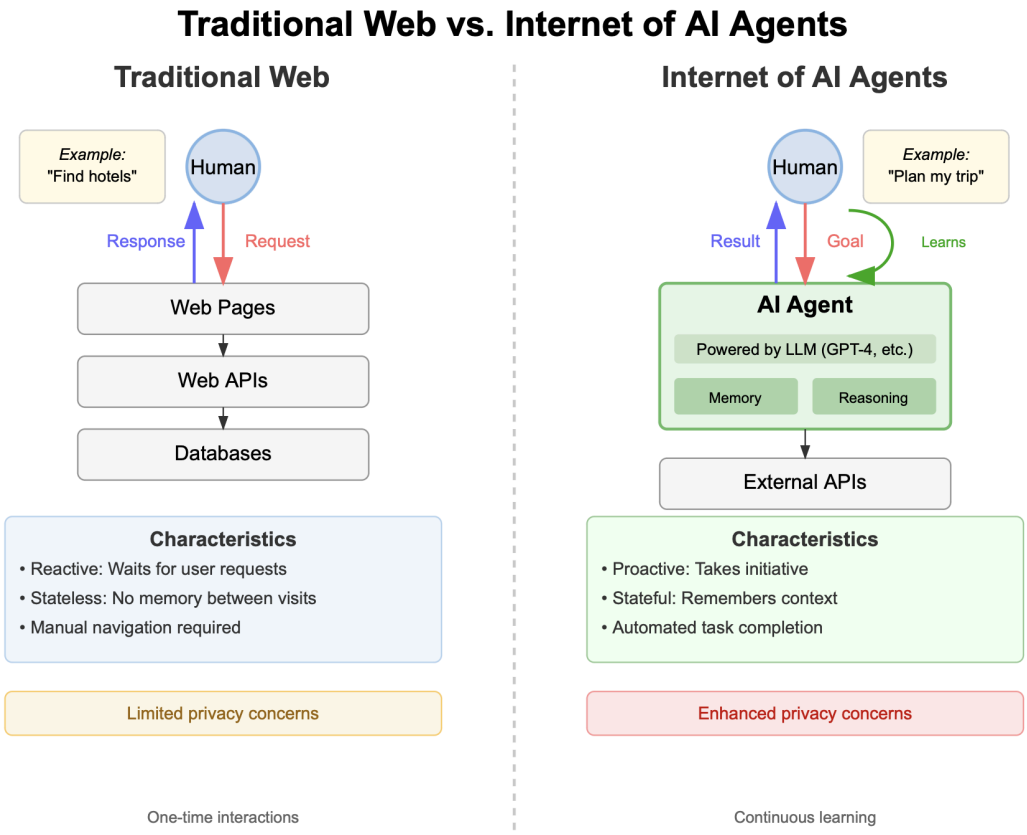

Figure 1: Traditional Web vs. Internet of AI Agents Architecture. The traditional web operates on a reactive request-response model where users manually navigate through web pages, APIs, and databases to find information. In contrast, the internet of AI agents introduces intelligent AI agents powered by Large Language Models (LLMs) that proactively interpret user goals, maintain memory of past interactions, and employ reasoning capabilities to autonomously complete complex tasks. While traditional web interactions are stateless and require explicit user requests, AI agents in the internet of AI agents continuously learn from interactions, remember context, and can take initiative to fulfill user intentions. This fundamental shift from reactive information retrieval to proactive task completion comes with enhanced privacy considerations due to the autonomous nature of data processing by AI agents.

A Primer on the WWW Architecture and Hierarchy

The current web architecture relies on several key components that work together to enable the global network we use daily.

The Domain Name System (DNS) serves as the web's phone book, translating human-readable domain names (like example.com) into machine-readable IP addresses. This hierarchical, distributed database system provides globally unique identifiers for websites, a hierarchical namespace structure (root, top-level domains, second-level domains), distributed management through multiple registrars, and resolution services with propagation times typically measured in hours.

The WHOIS Database complements DNS by providing metadata about domain ownership, including contact information for domain owners, registration and expiration dates, name server information, and limited verification of identity.

IP Addressing provides unique identifiers for devices connected to the internet. IPv4 uses 32-bit addresses, limiting the namespace to approximately 4.3 billion addresses, while IPv6 uses 128-bit addresses, theoretically allowing for 2^128 unique addresses — that's about a billion-billion-billion times the estimated number of grains of sand on Earth.

Certificate Authorities issue digital certificates that authenticate website identities and enable secure communication. They validate domain ownership, issue certificates with expiration dates, maintain certificate revocation lists, and operate at human-oriented speeds and verification levels.

This architecture has served the human-centric web remarkably well for decades but was designed with human timescales and interaction patterns in mind. The question now is whether this foundational architecture can evolve to support the emerging requirements of autonomous agents.

A Lesson from Dialup → Broadband

The transition from dialup to broadband internet provides valuable insights into how we might approach the shift to the internet of AI agents. When the internet was first commercialized, existing telephone infrastructure seemed like a natural fit - it already connected most homes and businesses. However, as internet usage evolved, fundamental limitations of dialup became apparent.

Why We Didn't Use Dialup for Internet Long-Term

Dialup's limitations revealed the importance of designing infrastructure for future needs rather than just current requirements. The connection model was fundamentally flawed for internet use, as dialup's circuit-switched, temporary connections were ill-suited for the persistent connectivity the internet would ultimately demand.

The addressing system also proved inadequate. Phone numbers weren't designed as internet endpoints, lacking the hierarchical structure needed for efficient routing. Additionally, the system suffered from one-way connections, where home machines could dial up the mainframes but not the other way around.

Scalability emerged as another critical limitation. The telephone system was designed for human-to-human voice communication, not for the massive data transfers that would become common with the growth of the internet.

How We Dealt with Known Unknowns

Several challenges were anticipated in the transition to broadband internet. Engineers knew dialup's 56 Kbps maximum would be insufficient for emerging applications, requiring significantly faster speeds. The need for always-on connectivity was recognized early, necessitating stateful and persistent connections rather than temporary dial-up sessions.

The architecture also had to evolve from point-to-point telephone calls to network models that could support many-to-one and one-to-many connections, enabling both client-server and peer-to-peer

architectures. Additionally, there was a shift from numerical phone numbers to domain names that could be easily remembered and shared, creating a human-readable namespace.

Finally, the system needed to accommodate the explosive growth of the web, requiring an addressing and discovery system that could scale to billions of web pages and resources, far beyond anything the telephone network had envisioned.

How We Prevented Unknown Unknowns

The move to packet-switched networks with TCP/IP created a flexible foundation that could accommodate unforeseen developments in several key ways. The open architecture of IP networking allowed for bandwidth-intensive applications no one could have predicted, such as video and streaming. The layered protocol design enabled security features to be added without disrupting the underlying infrastructure, leading to the development of HTTPS and modern security protocols. The addressing system provided the ability for a single IP address to host multiple services and applications, enabling the growth of SaaS and diverse endpoints. Finally, the flexible addressing scheme allowed for entirely new categories of connected devices, including mobile phones and IoT devices, demonstrating how a well-designed foundation can support innovations beyond its original conception.

This history demonstrates that successful infrastructure transitions require both addressing known challenges and building in flexibility for unforeseen developments. The agentic web will likely require similar foresight—particularly as billions of AI agents begin operating simultaneously across networks, creating unprecedented demands for coordination, resource allocation, and security that we cannot fully anticipate today.

The broadband transition illustrates a deeper truth: infrastructure does not just enable applications—it shapes what kinds of applications are even conceivable. Packet-switching, always-on connections, and scalable routing did not simply improve internet use; they redefined it. Similarly, the architecture we choose for agent registries will either constrain or unlock future capabilities and governance models for digital agents.

Possible Extended Use of Current DNS

Before examining entirely new architectures, it's worth considering how the current DNS system might be extended to accommodate autonomous agents:

Walled Gardens: Few IP Addresses but Many Endpoints

One approach would be to treat agents similarly to how web applications are currently managed: Each agent platform (like Salesforce, Microsoft CoPilots, or specialized agent hosting services) would maintain a range of IP addresses. Individual agents such as Delta Airlines would be identified by paths or subdomains (salesforce.com/DeltaAgent1 or DeltaAgent1.salesforce.com/). Internal routing would direct traffic to specific agents.

However, this approach faces several challenges. While it leverages existing infrastructure, it creates potential bottlenecks and single points of failure, as well as security and compliance issues. Each walled garden will have to agree to global protocols to talk to agents in another walled garden, which will add latency and performance variations. The walled garden hosts may throttle performance based on an opaque set of rules.

Privacy concerns also emerge with this model, as the walled garden host will have whole observability and monitoring access for any agent. Additionally, while this would allow users to initiate a call with DeltaAgent1, agents will not be able to call users.

Each Agent with an IPv4 Address

Alternatively, agents could be given their own IPv4 addresses, providing direct addressability for each agent, such as @DeltaAgent1 = x.y.z.w. This approach leverages existing DNS infrastructure but faces significant challenges. The system would be severely limited by the IPv4 address space (4.3 billion addresses, of which over 90% are already allocated according to IANA, with critical regions like APNIC and RIPE NCC having exhausted their allocations since 2011), making it inadequate for billions or trillions of agents.

Each Agent with IPv6 Address

IPv6 offers a much larger address space, with 2^128 possible addresses that could theoretically accommodate trillions of agents. This approach would allow directly addressable agents without intermediaries while maintaining compatibility with existing DNS systems. However, there are significant challenges: the registries will be massive with high latency, errors and propagation delays. Additionally, this requires full IPv6 adoption, which has been slow (global IPv6 adoption stands at approximately 44%, with significant variation across countries [5]).

Each approach has significant limitations when considering the potential scale and requirements of the internet AI of agents.

An Important Consideration: Search Path Configuration and Domain Boundaries

An essential capability that must be addressed is the concept of configurable search paths for agents, similar to how DNS operates today. Currently, DNS allows seamless movement between intranet and internet resources, automatically routing queries through organizational domains before public resolution. However, this paradigm doesn't exist for search - organizations cannot effectively intersect their internal search with public search, despite attempts with search appliances.

For the internet of AI agents to be viable in enterprise environments, agents must respect configurable search paths. Organizations need the ability to prioritize internal agent discovery and communication while maintaining controlled access to public domain agents. Additionally, robust traceability is required to monitor when agent interactions cross boundaries from private "walled gardens" to public domains, ensuring compliance and security policies are maintained. Without this functionality, enterprises will be unable to adopt agent-based architectures that require both internal and external agent collaboration.

WebPages vs Agents

To understand why current architectures may be insufficient, we must recognize fundamental differences between static web resources and autonomous agents:

Static Pages vs Agents as Programs

Web pages are primarily passive resources where content is served upon request, interactions are limited and predefined, changes are typically made by human administrators, and identity is relatively stable.

Agents, by contrast, are dynamic, evolving entities. They initiate actions and requests and can be ephemeral, appearing and disappearing in the thousands per day. They make autonomous decisions, may change behavior over time, and may create other agents. Additionally, they may move between hosting environments, require more complex identity verification, and require significant computational resources.

Potential Issues

These differences create several potential issues when trying to use current web infrastructure:

1. **Latency** - Agents may require much faster response times than human-oriented DNS resolution provides
2. **Security** - Autonomous agents introduce new attack vectors, including prompt injections and unauthorized API access, necessitating advanced security measures beyond traditional protocols
3. **Traceability** - The autonomous and potentially self-modifying nature of agents complicates tracking their decision-making processes, making accountability and compliance more challenging
4. **Verification** - Determining an agent's capabilities, trustworthiness, and ownership requires more sophisticated mechanisms than current certificate systems
5. **Scalability & Coordination:** The internet of AI Agents enables distributed coordination and adaptive task routing, offering resilience and scalability—whereas today's infrastructure struggles with isolated services and is prone to outages.
6. **Data & Application Models:** The classic web relies on rigid APIs and siloed data with limited interoperability, while agentic systems embrace semantic web standards, support dynamic interactions, and utilize machine-readable intents for seamless negotiation between agents.

These fundamental differences suggest that simply extending current systems may not be sufficient for the agentic web's needs. New aspects, such as the dynamic verification of behavior, traceable decision histories, and revocable privileges must be paired with behavioral metadata, reputational scores, or cryptographic attestations to allow systems and users to assess agent reliability in real-time.

Challenges in Scaling - Unique Crossover Points

As we consider the transition to the internet of AI agents, several scaling challenges emerge that represent critical "crossover points" where current systems may break down:

Known Unknowns

Several critical technical limitations emerge when considering the current internet infrastructure for agent management. First, the **number of unique IP addresses** presents a major constraint, as the current IPv4 space is already exhausted and IPv6 adoption remains uneven and slow.

DNS update propagation poses another significant challenge [3]. End‑user visibility is gated by **public recursive resolvers**: some consumer ISPs cache records aggressively or ignore low‑TTL settings, so worst‑case propagation across the full resolver ecosystem can still stretch to **24–48 hours** [4], which becomes unworkable for billions or trillions of agents with dynamic capabilities or locations. Agent directories might need near real-time updates, something the current system cannot provide.

New metadata requirements also present difficulties. Current WHOIS data is minimal and oriented toward human ownership, while agents will require rich metadata about capabilities, permissions, ownership, and other characteristics. Existing systems have no standard way to represent or query this information.

Finally, **certificate issuing and verification** creates additional bottlenecks. Current certificate authorities operate on human timescales (days to weeks) [7], but agent creation, modification, and verification may need to happen in seconds. Additionally, certificate revocation lists would become unmanageably large [8].

Unknown Unknowns

Several areas present challenges we can anticipate but not fully define. **Latency** requirements and **security** models raise fundamental questions about how quickly agents will need to discover and authenticate other agents, and what new attack vectors will emerge in agent-to-agent interactions.

Governance and **interoperability** present equally complex challenges. We must determine how standards will be established and enforced across potentially trillions of agents, and how agents from different frameworks and with different capabilities will communicate effectively.

Privacy concerns are particularly significant. The protocols and standards that will ensure robust data minimization and privacy in an agentic web remain undefined, raising open questions about how user data can be protected across billions of autonomous agents.

Data storage and **management** present emerging needs for privacy-preserving compute and federated data access, raising questions about the adoption of secure enclaves, homomorphic encryption, and other advanced storage models. Request handlers face a related challenge: handling billions of real-time agentic requests will require breakthroughs in scalable, distributed load balancing and intent-aware orchestration. Whether existing architectures can be extended or if entirely new coordination models are needed is still unknown.

Upgrade or Switch

Given these challenges, we face a fundamental decision: should we upgrade existing web infrastructure or switch to entirely new systems designed specifically for the internet of AI agents?

Upgrade Options

Several potential solutions build on existing infrastructure. First, **switching to IPv6** would leverage its massive address space while maintaining compatibility with existing DNS infrastructure, providing a gradual transition path from current systems.

Updating WHOIS metadata represents another evolutionary approach. This would involve extending existing WHOIS databases with agent-specific fields, adding capability descriptions, trust metrics, and other agent-specific data, while standardizing query methods for agent-specific attributes.

Automated certificate methods could address verification needs by developing faster, more automated certificate issuance systems. This would include creating agent-specific trust metrics and verification standards, as well as implementing near real-time certificate revocation mechanisms.

DNS automation offers a final path forward, dramatically improving DNS propagation times, implementing agent-specific DNS record types, and creating specialized agent directory services within the existing DNS hierarchy.

Switch Options

More revolutionary approaches could fundamentally reimagine agent addressing. Creating a **new addressing system** for agents would involve developing a parallel addressing system specifically for agents, designed for agent-specific requirements from the ground up, potentially implementing entirely new resolution protocols.

Decentralized identity systems offer another paradigm shift, implementing DID (Decentralized Identifier) based systems that allow agents to maintain self-sovereign identity and enable direct agent-to-agent verification without central authorities [6].

Hybrid registry systems could balance centralized and decentralized approaches by maintaining centralized registries for critical or high-trust agents while using decentralized mechanisms for the long tail of specialized agents, creating bridge protocols between different registry systems.

Capability-based addressing represents perhaps the most radical departure from current systems. This approach would address agents by their capabilities rather than static identifiers, enable dynamic discovery based on required functions, and implement semantic addressing systems.

Comparative Analysis

Factor	Upgrade	Switch
Number of IP Addresses	IPv6 with 128-bit addresses	New IP schema designed specifically for agents
Update and Propagation	Automate existing DNS update machinery	Use decentralized mechanisms (DID and SSI)
Certificate Management	Enhance current system for faster issuance	Implement DIDs with built-in verification
New Metadata	Add new fields to existing WHOIS records	Create agent-specific directories with rich metadata
Capability Discovery	Limited extensions to DNS	Native capability-based addressing
Implementation Timeline	Faster initial deployment	More robust long-term solution
Backward Compatibility	High	Limited
Governance	Extend existing ICANN model	Requires new governance frameworks

Conclusion

The transition to the internet of AI agents represents a fundamental shift in how we interact with digital systems, comparable to the shift from dialup to broadband internet. While upgrading existing systems offers a path of least resistance and backwards compatibility, the unique requirements of autonomous agents suggest that entirely new registry architectures may ultimately be necessary.

The history of technology transitions suggests that hybrid approaches often emerge during periods of rapid change. We may see centralized registries for critical agents alongside decentralized systems for specialized agents, with bridge protocols enabling interoperability.

What's clear is that this infrastructure question must be addressed proactively rather than reactively. The decisions made today about agent registry architecture will shape the capabilities, security, and accessibility of the internet of AI agents for decades to come. By learning from past transitions and anticipating future needs, we can build infrastructure that not only addresses current requirements but remains flexible enough to accommodate the unknown unknowns that will inevitably emerge as autonomous agents become an integral part of our digital landscape.

Rather than simply extending human-oriented web infrastructure, we have an opportunity to design systems specifically for agent-to-agent interactions, potentially unlocking entirely new categories of

applications and services. Whether through upgrade or switch – or most likely, some combination of both – the registry architecture for the internet of AI agents will be a critical foundation for the next era of digital innovation. By acknowledging that infrastructure defines both capabilities and accountability, we can create an internet of AI agents that is not only scalable and performant, but also responsible and resilient.

References

1. DataReportal & Kepios. (2025, April). *Digital 2025 April Global Statshot Report*. https://datareportal.com DataReportal – Global Digital Insights
2. Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th ACM Symposium on User Interface Software and Technology (UIST '23)*. https://doi.org/10.1145/3586183.3606763 arXiv
3. Gao, Z., & Venkataramani, A. (2019, April). Measuring update performance and consistency anomalies in managed DNS services. In IEEE INFOCOM 2019-IEEE conference on computer communications (pp. 2206-2214). IEEE.
4. DNS Made Easy. (2025, March 25). DNS propagation: Why doesn't my domain work? DNS Made Easy. https://dnsmadeeasy.com/resources/dns-propagation-why-doesnt-my-domain-work
5. Google. (2025, May 6). IPv6 adoption statistics. Retrieved May 13, 2025, from https://www.google.com/intl/en/ipv6/
6. World Wide Web Consortium. (2025, Apr.). *Decentralized identifiers (DIDs) v1.1* (W3C Recommendation).
7. Chamola, S. (2023, September 22). EV SSL certificates: Pros & cons. Encryption Consulting. https://www.encryptionconsulting.com/ev-ssl-certificates-pros-cons/
8. Barnes, R., Hoffman-Andrews, J., McCarney, D., & Kasten, J. (2019). Automatic Certificate Management Environment (ACME) (RFC 8555). Internet Engineering Task Force. https://doi.org/10.17487/RFC8555