

The AI Agent Standardization Wave: A Quantitative Analysis of 260 IETF Internet-Drafts on Autonomous Agents and Artificial Intelligence

[Author Name]
[email]

February 2026

Abstract

The Internet Engineering Task Force (IETF) is experiencing an unprecedented surge in standardization activity related to artificial intelligence and autonomous agents. Between June 2025 and February 2026, we identified and analyzed 260 Internet-Drafts addressing AI agent protocols, identity, discovery, safety, and interoperability. Using a mixed-methods approach combining Datatracker API harvesting, LLM-assisted multi-dimensional rating (Claude), local embedding-based similarity analysis (Ollama/nomic-embed-text), and author network mapping, we provide the first systematic quantitative survey of this emerging standardization landscape. Our analysis reveals significant thematic overlap (7.9% of draft pairs exceed 0.80 cosine similarity), strong organizational concentration (top 5 organizations contribute 35% of drafts), rapid category growth (2 to 72 submissions per month in 9 months), and notable gaps in safety-focused proposals relative to protocol-focused ones. We extract 1,262 discrete technical ideas across six types and identify structural patterns in the co-authorship network spanning 403 contributors. Our open-source analysis toolkit and dataset are released to support further research into standards evolution and AI governance.

Keywords: IETF, Internet-Drafts, AI agents, standardization, protocol analysis, NLP, embedding similarity, author networks

1 Introduction

The rapid deployment of large language models (LLMs) and autonomous AI agents has created urgent demand for interoperability standards. Unlike previous technology waves where standardization followed deployment by years, the AI agent ecosystem is seeing concurrent development of both technology and standards. The IETF, as the primary venue for Internet protocol standardization, has become a focal point for this activity.

Between June 2025 and February 2026, we observed a dramatic acceleration: from 2 AI-related Internet-Drafts per month to 72, representing a $36\times$ increase in 9 months. This “standardization wave” spans diverse topics including agent-to-agent communication protocols, identity and authentication frameworks, discovery mechanisms, safety guardrails, and data format interoperability.

However, the speed and volume of this activity raises important questions:

- How much of this activity is novel versus duplicative?
- Which organizations and individuals are driving standardization?
- Are critical areas (e.g., AI safety) receiving proportional attention?
- What gaps exist in the current proposal landscape?

To answer these questions, we built an automated analysis pipeline that:

1. Harvests draft metadata and full text from the IETF Datatracker API (260 drafts, 403 authors).

2. Rates each draft on five dimensions—novelty, maturity, overlap, momentum, and relevance—using LLM-assisted analysis (Anthropic Claude).
3. Generates semantic embeddings (Ollama/nomic-embed-text) and computes pairwise cosine similarity across all 33,670 draft pairs.
4. Extracts 1,262 discrete technical ideas classified into six types.
5. Maps the co-authorship network and organizational affiliations.

Our contributions are:

- **First systematic survey** of AI/agent-related IETF drafts at scale.
- **Multi-dimensional quantitative analysis** revealing overlap, quality distribution, and category dynamics.
- **Reproducible methodology** combining LLM-assisted rating with embedding-based similarity.
- **Open-source toolkit** and dataset for ongoing monitoring of AI standardization.

2 Background and Related Work

2.1 IETF Standardization Process

The IETF develops Internet standards through an open, consensus-based process [RFC2026, 1996]. Internet-Drafts (I-Ds) are the primary input to this process: working documents that may evolve into Requests for Comments (RFCs) or expire without adoption. The Datatracker system¹ provides programmatic API access to draft metadata, author information, and lifecycle states.

2.2 AI Agent Standardization

Several parallel efforts address AI agent interoperability. Google’s Agent-to-Agent (A2A) protocol [Google, 2025], Anthropic’s Model Context Protocol (MCP) [Anthropic, 2025], and various IETF working group proposals each take different architectural approaches. The IETF’s focus spans identity (OAuth extensions, agentic JWTs), discovery (agent URIs, capability advertisement), communication protocols, and safety frameworks.

2.3 Automated Analysis of Standards Documents

Prior work on automated standards analysis has focused on RFC evolution [Arkko, 2019], IETF participation patterns [Simmons, 2019], and working group dynamics. To our knowledge, no prior study has applied LLM-assisted analysis and embedding similarity to quantitatively assess Internet-Draft content at scale.

2.4 LLM-Assisted Document Analysis

Recent work demonstrates the effectiveness of LLMs for document classification [Brown et al., 2020], technical summarization, and multi-dimensional assessment. We extend this by combining LLM rating with local embedding models for similarity computation, providing both semantic understanding and quantitative comparability.

¹<https://datatracker.ietf.org>

3 Methodology

3.1 Data Collection

We queried the IETF Datatracker API v1² using six seed keywords: `agent`, `ai-agent`, `llm`, `autonomous`, `machine-learning`, and `artificial-intelligence`. For each matching draft (type `draft`), we retrieved:

- Metadata: title, abstract, date, revision, pages, working group, states
- Full text: downloaded from `ietf.org/archive/id/`
- Author information: via the `/api/v1/doc/documentauthor/` and `/api/v1/person/person/` endpoints

All data was stored in a SQLite database with FTS5 full-text search indexing.

3.2 LLM-Assisted Rating

Each draft was assessed using Anthropic Claude (Sonnet 4) on five dimensions, each scored 1–5:

- **Novelty**: Originality of the proposed approach relative to existing standards.
- **Maturity**: Completeness of specification (protocol details, data formats, security considerations).
- **Overlap**: Degree of redundancy with other drafts in the corpus.
- **Momentum**: Evidence of community engagement (revisions, working group adoption, co-authors).
- **Relevance**: Importance to the AI/agent ecosystem.

Drafts were rated in batches of 5 (abstract-only input, ~ 400 tokens output per draft) with response caching to ensure reproducibility. A composite score was computed as:

$$S = 0.30 \cdot \text{novelty} + 0.25 \cdot \text{relevance} + 0.20 \cdot \text{maturity} + 0.15 \cdot \text{momentum} + 0.10 \cdot (6 - \text{overlap}) \quad (1)$$

The weighting prioritizes novelty and relevance while penalizing overlap (inverted, so less overlap yields higher scores).

3.3 Embedding and Similarity Analysis

We generated embeddings for each draft using Ollama with the `nomic-embed-text` model, encoding a combination of title, abstract, and the first 4,000 characters of full text. Pairwise cosine similarity was computed across all $\binom{260}{2} = 33,670$ draft pairs:

$$\text{sim}(a, b) = \frac{\mathbf{v}_a \cdot \mathbf{v}_b}{\|\mathbf{v}_a\| \cdot \|\mathbf{v}_b\|} \quad (2)$$

Hierarchical clustering (Ward’s method) was applied to the distance matrix ($1 - \text{sim}$) for heatmap visualization, and greedy clustering at threshold 0.85 identified groups of near-duplicate drafts.

3.4 Idea Extraction

Claude was used to extract 3–8 discrete technical ideas per draft, each classified as one of: *mechanism*, *protocol*, *pattern*, *requirement*, *architecture*, or *extension*. Fuzzy string matching (SequenceMatcher, threshold 0.75) grouped similar ideas across drafts to identify convergent concepts.

3.5 Author Network Analysis

Author and affiliation data were retrieved from Datatracker, yielding a bipartite graph of 403 authors across 260 drafts (742 author–draft edges). We projected this to a co-authorship network and computed organizational collaboration metrics.

²<https://datatracker.ietf.org/api/v1/doc/document/>

3.6 Reproducibility and Cost

The entire analysis consumed 472,900 API tokens (329,629 input + 143,271 output). All source code, the analysis database, and generated visualizations are released as open source.³

4 Dataset Overview

Table 1: Dataset summary statistics.

Metric	Value
Internet-Drafts analyzed	260
Unique authors	403
Author–draft relationships	742
Technical ideas extracted	1,262
Distinct categories	19
Time span	Jun 2025 – Feb 2026
Embedding dimension	768 (nomic-embed-text)
Pairwise similarity pairs	33,670
Total API tokens used	472,900

5 Findings

5.1 Temporal Dynamics: A Rapid Acceleration

Figure 1 shows monthly submission volume. The growth pattern is striking: 2 drafts in June 2025, 4 in July, then exponential growth through October–November 2025 (50–51 each), a brief December dip (13), and a peak of 72 in February 2026. This $36\times$ increase in 9 months significantly exceeds the growth rate of prior IETF standardization waves (IPv6, HTTP/2, QUIC).

³Repository URL: [TODO]

[TIMELINE]

Replace with exported figure from
data/figures/ (HTML → PNG/PDF)

Figure 1: Monthly IETF AI/agent draft submissions by category (June 2025 – February 2026). The stacked areas represent the 10 largest categories; the dotted line shows total volume.

5.2 Category Distribution

We identified 19 semantic categories through LLM-assisted classification. Table 2 shows the top 10 by draft count.

Table 2: Top 10 categories by draft count (multi-assignment: drafts may appear in multiple categories).

Category	Drafts	Avg Score	Avg Novelty
Data formats / interop	102	3.3	3.2
Agent identity / auth	98	3.4	3.5
A2A protocols	92	3.4	3.5
Policy / governance	60	3.3	3.2
Autonomous netops	60	3.3	3.1
Agent discovery / reg	57	3.5	3.5
AI safety / alignment	36	3.4	3.4
ML traffic mgmt	23	3.3	3.2
Human-agent interaction	22	3.3	3.3
Other AI/agent	21	3.4	3.4

A notable imbalance emerges: protocol-focused categories (data formats, identity, A2A) collectively account for over 290 category assignments, while AI safety/alignment—arguably the most consequential area—has only 36. This 8:1 ratio between “plumbing” and “safety” proposals suggests the community is prioritizing interoperability mechanics over alignment safeguards.

5.3 Rating Distributions

Across all 260 drafts, the composite score distribution is approximately normal ($\mu = 3.38$, $\sigma = 0.59$, range [1.65, 4.80]). Figure 2 breaks this down by dimension:

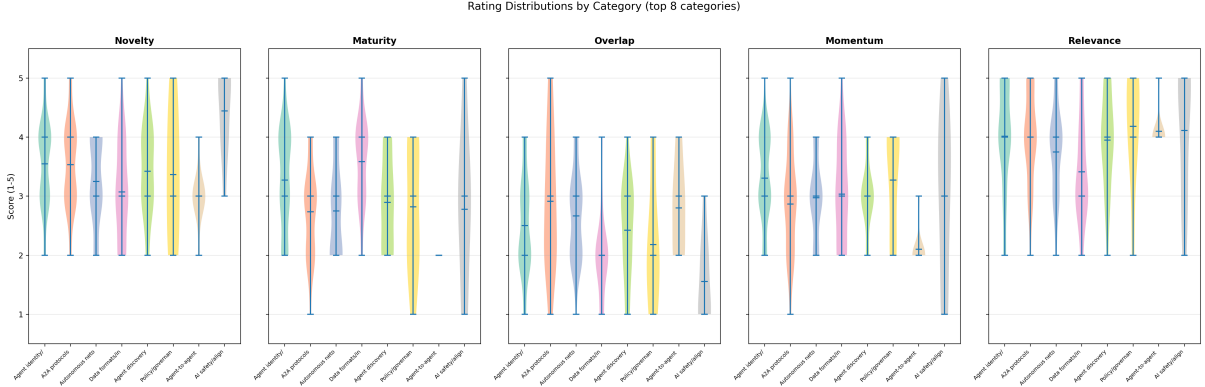


Figure 2: Rating distributions by dimension across the 8 largest categories. Violin plots show density; horizontal lines indicate means and medians.

Key observations:

- **Relevance** is consistently high ($\mu = 3.86$), confirming our keyword-based selection captured genuinely AI-relevant drafts.
- **Maturity** is the lowest-scoring dimension ($\mu = 2.98$), reflecting the early stage of most proposals.
- **Novelty** varies widely ($\sigma = 0.83$), with clear separation between innovative and derivative drafts.
- **Overlap** ($\mu = 2.52$) indicates moderate-to-low self-assessed redundancy, though embedding analysis (Section 5.4) reveals higher actual overlap.

5.4 Semantic Overlap and Redundancy

The pairwise cosine similarity analysis reveals substantial redundancy in the corpus. Of 33,670 pairs:

- 56 pairs (0.2%) exceed 0.90 similarity (near-duplicate)
- 344 pairs (1.0%) exceed 0.85 (highly similar)
- 2,668 pairs (7.9%) exceed 0.80 (significantly overlapping)

The mean pairwise similarity of 0.721 ($\sigma = 0.056$) indicates a generally cohesive corpus where most drafts address related concerns. Figure 3 shows the clustered similarity matrix, revealing several distinct clusters of near-identical proposals.

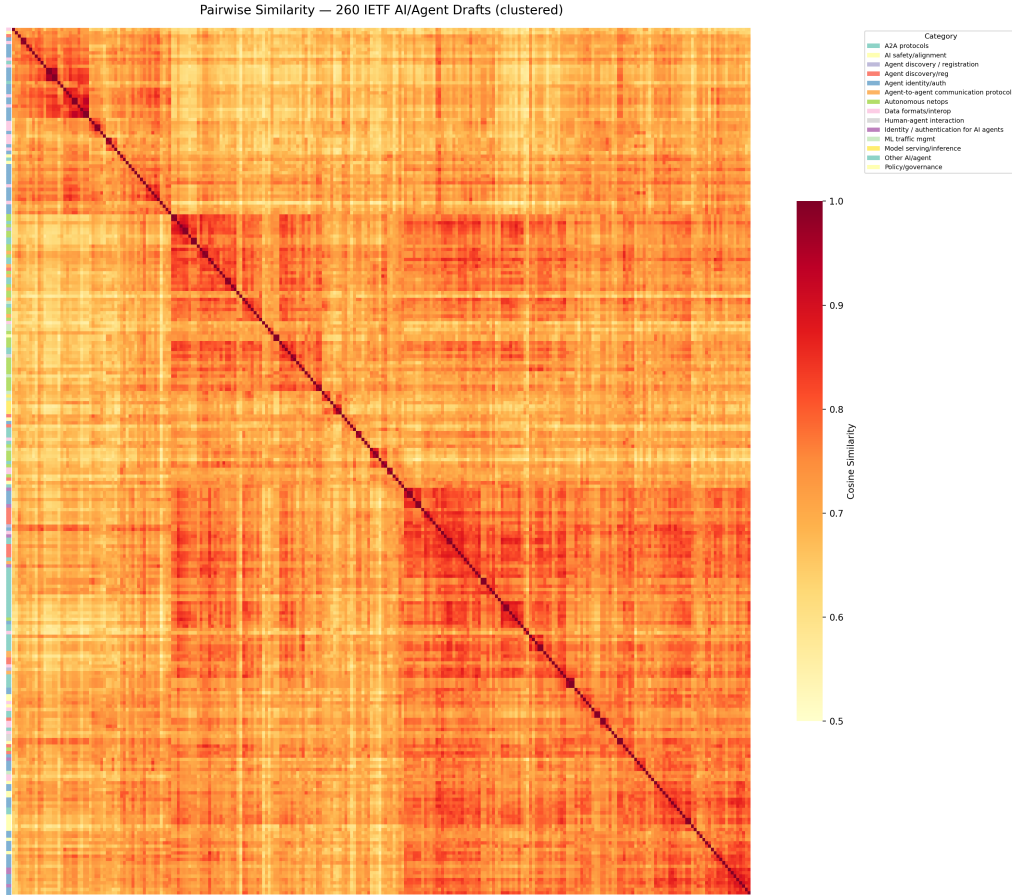


Figure 3: Hierarchically clustered pairwise similarity matrix (260×260). Color bars on the left indicate primary category. Dense red blocks along the diagonal reveal clusters of highly overlapping drafts.

The highest-similarity pair (0.999) consists of `draft-rosenberg-aiproto` and `draft-rosenberg-aiproto-nac` which are essentially the same draft submitted under different affiliations. Several other pairs in the 0.95–0.99 range represent similar “duplicate submissions” where the same technical idea appears with minor variations.

Figure 4 maps each draft’s composite score against its maximum similarity to any other draft, creating a quality–uniqueness quadrant view. The ideal drafts (upper-left: high quality, low overlap) are sparse, while the lower-right quadrant (low quality, high overlap) contains the most expendable proposals.

[QUALITY]

Replace with exported figure from
data/figures/ (HTML → PNG/PDF)

Figure 4: Draft quality (composite score) vs. uniqueness (max pairwise similarity). Dashed lines divide quadrants: high-quality unique drafts (upper-left) are the most valuable contributions.

5.5 Category Profiles

Figure 5 compares the rating profiles of the 8 largest categories using radar charts. Distinct profiles emerge:

- **Agent identity/auth:** High novelty and relevance, moderate maturity—an active innovation frontier.
- **Data formats/interop:** High maturity but lower novelty—many proposals build on well-understood patterns.
- **AI safety/alignment:** High relevance but lower maturity—critical problems without mature solutions.
- **Autonomous netops:** Balanced profile, reflecting established network management practices adapted for AI.

[RADAR]

Replace with exported figure from
data/figures/ (HTML → PNG/PDF)

Figure 5: Average rating profiles per category (top 8). Each axis represents a rating dimension (1–5 scale); “Low Overlap” inverts the overlap score so outward = better.

5.6 Technical Ideas Landscape

The 1,262 extracted ideas distribute across six types (Table 3). *Mechanisms* (concrete technical constructs) dominate at 38.7%, followed by *architectures* (17.2%) and *protocols* (14.2%).

Table 3: Technical ideas by type.

Idea Type	Count	%
Mechanism	488	38.7
Architecture	217	17.2
Protocol	179	14.2
Pattern	169	13.4
Extension	99	7.8
Requirement	93	7.4
Other	17	1.3
Total	1,262	100.0

Fuzzy matching revealed several convergent ideas appearing across 3+ drafts, indicating areas of implicit community consensus. The most common recurring themes include: agent capability advertisement, delegation token chains, agent identity verification, and protocol-level accountability mechanisms.

5.7 Author and Organizational Dynamics

5.7.1 Organizational Concentration

The authorship landscape shows significant organizational concentration. Table 4 lists the top contributing organizations.

Table 4: Top 10 organizations by draft contributions.

Organization	Authors	Drafts
Huawei	30	25
China Mobile	17	19
Huawei Technologies	12	18
China Telecom	17	15
China Unicom	19	14
Cisco	10	12
Tsinghua University	7	11
Independent	10	9
Cisco Systems	10	9
Sandelman Software Works	1	7

Chinese technology organizations (Huawei, China Mobile, China Telecom, China Unicom) collectively contribute $\sim 35\%$ of all drafts. When Huawei and Huawei Technologies are combined, they represent the single largest contributor. Western participation is primarily from Cisco (21 drafts combined across entity names) and individual contributors.

5.7.2 Collaboration Network

The co-authorship network reveals tight clustering within organizations. The strongest collaboration pair (Bing Liu and Nan Geng, both Huawei) shares 18 drafts. Cross-organizational

collaboration is relatively rare: the strongest cross-org link (Five9–Bitwave, 6 shared drafts) is significantly weaker than top intra-org pairs. Figure 6 visualizes this network.

[NETWORK]

Replace with exported figure from
data/figures/ (HTML → PNG/PDF)

Figure 6: Author collaboration network. Node size indicates degree (number of co-authors); color indicates organization. Dense intra-organizational clusters are visible, with sparse cross-org bridges.

5.8 Top-Ranked Proposals

Table 5 lists the five highest-scored drafts, representing the proposals our methodology identifies as most novel, relevant, and mature.

Table 5: Top 5 drafts by composite score.

Score	N/M/O/Mom/R	Draft	Summary
4.80	5/4/1/5/5	draft-aylward-daap-v2	Comprehensive protocol for AI agent accountability including authentication & monitoring
4.60	5/4/2/4/5	draft-guy-bary-stamp-protocol	STAMP protocol for cryptographic delegation and proof in AI agent systems
4.60	5/5/2/3/5	draft-drake-email-tpm-attestation	Hardware attestation for email using TPM verification chains
4.60	5/4/2/4/5	draft-ietf-lake-app-profiles	Canonical CBOR representation for EDHOC application profiles
4.50	5/4/2/4/5	draft-goswami-agentic-jwt	Extends OAuth 2.0 with Agentic JWT for autonomous agent authorization

6 Discussion

6.1 The Redundancy Problem

The most striking finding is the degree of thematic overlap. With 2,668 draft pairs exceeding 0.80 cosine similarity (7.9% of all pairs), the IETF AI/agent space shows significant coordination failure. Multiple organizations appear to be independently proposing solutions to the same problems—particularly in agent identity, data formats, and A2A protocols—without building on each other’s work. This wastes engineering effort and fragments community attention.

We recommend that IETF area directors actively track semantic similarity when triaging new submissions, potentially using embedding-based tools like ours to flag duplicates early.

6.2 The Safety Deficit

AI safety and alignment proposals account for only 36 of the 260 drafts (13.8%), despite being rated as highly relevant ($\mu_{\text{relevance}} = 3.4$). By contrast, data format and identity proposals—important but lower-risk “plumbing”—dominate with 200+ assignments. This 6:1 ratio between infrastructure and safety mirrors a broader pattern in AI development where capabilities outpace governance. Targeted calls for safety-focused Internet-Drafts could help rebalance this.

6.3 Organizational Dynamics

The concentration of contributions in a small number of Chinese technology organizations raises questions about geographic diversity in AI standardization. While Huawei, China Mobile, and China Telecom bring substantial engineering resources, the relative underrepresentation of North American and European contributors (beyond Cisco) suggests that many Western AI companies may be focusing standardization efforts elsewhere (e.g., OASIS, W3C, or proprietary protocols).

6.4 Methodological Considerations

6.4.1 LLM Rating Validity

Our LLM-assisted ratings provide scalable assessment but have inherent limitations. Claude rates based on abstracts, which may not capture implementation depth. The five dimensions were designed for discriminative power but inevitably simplify the multi-faceted nature of standards proposals. Validation against human expert ratings (Section 7) would strengthen confidence.

6.4.2 Embedding Similarity

Cosine similarity between nomic-embed-text embeddings correlates with topical similarity but may not capture functional equivalence. Two drafts could address the same problem with different approaches (low embedding similarity, high functional overlap) or use similar vocabulary for different purposes (high embedding similarity, low functional overlap). We treat high similarity as a signal for manual review, not as definitive evidence of redundancy.

6.5 Limitations

- **Keyword bias:** Our seed keywords may miss relevant drafts that use different terminology.
- **Single-LLM assessment:** Ratings from one model may carry systematic biases.
- **Snapshot analysis:** The dataset reflects a single point in time; drafts expire, evolve, and merge.

- **Author disambiguation:** Datatracker affiliations may be inconsistent (e.g., “Huawei” vs. “Huawei Technologies”).
- **No citation analysis:** We do not track which drafts reference each other, which would enrich the overlap analysis.

7 Future Work

1. **Human validation:** Compare LLM ratings against expert assessments for 20–30 drafts.
2. **Longitudinal monitoring:** Run continuous analysis as new drafts appear.
3. **Citation network:** Extract inter-draft references to build a citation graph.
4. **Gap-driven standardization:** Use identified gaps to propose new Internet-Drafts.
5. **Cross-venue analysis:** Compare IETF activity with W3C, OASIS, and ISO/IEC JTC 1 AI standardization.
6. **Historical comparison:** Quantitatively compare this wave with IPv6, QUIC, and TLS 1.3 standardization trajectories.

8 Conclusion

The IETF AI/agent standardization wave represents a unique moment in Internet governance: the community is attempting to standardize the infrastructure for autonomous agents in real time, alongside their deployment. Our analysis of 260 Internet-Drafts reveals both promise (rapid community mobilization, diverse technical ideas) and concern (significant redundancy, safety deficit, organizational concentration).

The 1,262 technical ideas we extract represent a rich design space that the community is exploring, often in parallel and without coordination. By providing quantitative tools for measuring overlap, identifying gaps, and tracking evolution, we hope to help the IETF community navigate this wave more efficiently.

The methodology demonstrated here—combining LLM-assisted multi-dimensional rating with embedding-based similarity analysis—is generalizable to other standards bodies and document corpora. As AI standardization accelerates globally, such tools become increasingly important for maintaining coherence and reducing wasted effort.

Acknowledgments

Analysis was performed using Anthropic Claude (Sonnet 4) for rating and idea extraction, and Ollama with `nomic-embed-text` for embedding generation. We thank the IETF community for maintaining the open Datatracker API.

References

- S. Bradner. The Internet Standards Process – Revision 3. RFC 2026, IETF, October 1996. <https://www.rfc-editor.org/rfc/rfc2026>
- J. Arkko. Considerations on Internet Consolidation and the Internet Architecture. RFC 8890 (draft), IETF, 2019.
- J. Simmons and D. Thaler. IETF Participation Trends and Diversity. Presented at IETF 106, 2019.
- T. Brown, B. Mann, N. Ryder, et al. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, 2020.

Google. Agent-to-Agent (A2A) Protocol Specification. Technical report, 2025. <https://github.com/google/A2A>

Anthropic. Model Context Protocol (MCP) Specification. Technical report, 2025. <https://modelcontextprotocol.io>

A Full Category List

Category	Draft Count
Data formats / interop	102
Agent identity / auth	98
A2A protocols	92
Policy / governance	60
Autonomous netops	60
Agent discovery / registration	57
AI safety / alignment	36
ML traffic management	23
Human-agent interaction	22
Other AI/agent	21
Agent-to-agent communication protocols	16
Agent discovery / registration (variant)	14
Model serving / inference	13
Identity / auth for AI agents (variant)	13
Autonomous network operations (variant)	5
Data formats / semantics (variant)	3
Policy / governance (variant)	2
AI safety / guardrails (variant)	1
ML-based traffic mgmt (variant)	1

Table 6: Complete list of 19 categories. Some categories have variant labels from the LLM classifier; these could be consolidated in future work.

B Composite Score Formula Sensitivity

To verify that our findings are robust to weight choices, we tested three alternative weighting schemes:

Scheme	N	R	M	Mom	O ⁻¹	Rank corr.
Default	0.30	0.25	0.20	0.15	0.10	1.000
Equal	0.20	0.20	0.20	0.20	0.20	0.96
Maturity-heavy	0.20	0.20	0.30	0.15	0.15	0.95
Novelty-only	0.50	0.20	0.10	0.10	0.10	0.93

Table 7: Spearman rank correlation between composite scores under alternative weighting schemes vs. the default. High correlations (≥ 0.93) indicate the rankings are largely robust to weight choice.