AIFeed: Verifiable Content Permissions and Efficient Agent Delivery for the AI Web

AIFeed Protocol Contributors*

Draft — September 16, 2026

Abstract

AI systems now consume more web content than people do, and the plain-text preferences in `robots.txt` do not hold them back: one vendor logged 1.9 billion crawls that ignored robots rules in a single half-year, and one web-only measurement put the crawl-to-referral ratio of a major AI provider at 70,900:1. Preference and licensing signals exist, but they are not attributable to a domain, cannot be revoked, and do nothing about the cost of repeated consumption. We describe AIFeed, an open trust layer built around a signed manifest of machine-readable permissions, anchored in DNS and checked against a multi-signature revocation registry. Two content profiles ride on the same signed bytes: a native markdown format (AIFeed Markdown) and a compatibility profile for the external MAKO draft. Three properties, taken together, distinguish the design from the signals we surveyed: signed provenance of permissions, per-page binding with restrict-only overrides, and a digest-bearing delta index that lets an agent skip pages that have not changed. We measure the system on committed artifacts. Converting a 60-page corpus to the markdown profiles cuts transferred bytes by 68.8% (95.7% for delta consumption), and a loopback enforcement harness with four client profiles records publisher savings of 55.2% of bytes and 56.2% of CPU, AI-side savings of 54.8% (72.9% for the compliant client), 14 of 18 unchanged pages skipped, and signature verification at 0.70 ms per page. The specification is exercised by 25 manifest, 39 MAKO, and 11 AIFeed Markdown vectors under independent JavaScript and Python verifiers, plus differential PHP fixtures and an end-to-end WordPress deployment. We also report results that do not flatter the design: without adoption and enforcement there are no savings at all; vendor claims of up to 94% token reduction rely on summarization our converter does not perform; origin-plus-DNS compromise is invisible on first contact; and the compatibility profile depends on a third-party draft. No external cryptographic review or live pilot exists yet.

*Contact: `contact@aifeed.md` (submission contact; to be confirmed). Artifacts: `https://github.com/denyni/aifeed-protocol`.

1 Introduction

Web content is increasingly consumed by automated agents rather than human visitors. Cloudflare's radar puts AI bots at an average of 4.2% of HTML requests in 2025 (6.4% at peak), with automated traffic approaching half of all HTML requests and "user action" agent crawls growing more than 15-fold over the same period [23]. The publisher-facing asymmetry is worse than those aggregates suggest: one web-only measurement found a crawl-to-referral ratio of 70,900:1 for a major AI provider [21], and a single vendor recorded 22 billion AI scrapes in a half-year, 1.9 billion of which ignored `robots.txt` [52].

The ecosystem is not empty. RFC 9309 and the IETF AIPREF drafts cover robotic preferences; Cloudflare's Content Signals, RSL, and the W3C TDM Reservation Protocol each express some flavor of permission or license [1, 2, 6, 7, 20, 35]; `llms.txt` indexes a site for language models [5]; and on the other side of the exchange, Web Bot Auth is putting signed agent requests into production [4, 19]. What is missing is twofold. A publisher's declaration is not attributable: any intermediary can alter it, and nothing standard lets the publisher revoke it. And even a declaration that is respected leaves the mechanical cost untouched, because agents keep pulling unchanged pages as browser-shaped HTML.

We make four contributions:

1. **Design.** AIFeed v0.1: an Ed25519-signed, JCS-canonicalized manifest at `/.well-known/ai.json`, cryptographically anchored to the domain through a DNS TXT record, servable offline, with a multi-signature revocation registry and bounded staleness semantics.

2. **Content profiles.** AIFeed v0.2 adds signed per-page content delivery: the native AIFeed Markdown profile (`text/aifeed+markdown`) with first-class permission, asset, and triage metadata; the MAKO compatibility profile; restrict-only per-page overrides; and a digest-bearing delta index with a site resume and per-entry triage fields.

3. **Implementation.** A zero-dependency reference stack: CLI, JavaScript SDK, independent Python verifier, WordPress publisher plugin, static-site builder, and server adapters for nginx, Caddy, Apache, Node, Next.js, PHP, Python ASGI, and Go.

4. **Evaluation and reproducibility.** Measured conversion, delta, and enforcement results with committed artifacts, conformance vectors, fuzzing, and differential cross-language testing; unfavorable results are reported alongside them.

2 Background and Related Work

Preference and permission signals. Robots preferences were standardized in RFC 9309 [35], and the IETF AIPREF working group is now assembling a vocabulary for AI usage preferences together with an attachment mechanism for HTTP responses [1, 2]. Industry has moved faster: Cloudflare's Content Signals Policy expresses search, AI-input, and AI-training preferences and reports adoption on more than 3.8M domains through managed robots files [20], with pay-per-crawl billing offered at the same edge [22]; RSL adds licensing and compensation terms [6]; the W3C TDM Reservation Protocol handles text-and-data-mining reservations [7]; and `llms.txt` gives language models a site-level content index [5]. What none of these provides is attribution of the declaration to the domain, a revocation path, or verification that does not depend on the transport channel.

Agent-side authentication. Web Bot Auth defines HTTP message signatures for automated traffic, built on RFC 9421 and Ed25519 [3, 4, 14, 33]. Production deployments include a major assistant signing its agent requests and a platform verifying them [19]. AIFeed uses the same primitives in the opposite direction—the publisher signs—so the two directions compose rather than compete.

Content formats for agents. MAKO defines a per-page markdown format with YAML frontmatter, semantic links, declared actions, and experimental compact embeddings, and explicitly leaves authenticity verification out of scope [31]; vendor "markdown for agents" conversion exists at the CDN edge. Several projects from 2026 occupy adjacent ground. `ai-policy.json` gathers permission declarations at a well-known URL, though without signatures, anchoring, or revocation [42]; `agents.txt` puts identity, terms, and endpoints in a root file [13]; CrawlWall enforces crawler policy at the edge and keeps a signed audit ledger with receipts [32]; and `terms.txt` goes furthest, specifying per-path, per-purpose terms with a signed exchange, delegation tokens, and payment negotiation [18]. On the academic side, `ai.txt` proposes a domain-specific language for guiding AI interactions [40]. As of our inspection (2026-09-15), we did not find a single project that combines publisher-signed permissions with a DNS anchor and multi-signature revocation, a native agent format plus a compatibility profile over one signed payload, and a digest-bearing delta index with verifiable per-page triage. Our claim is to that combination, not to having invented any of its parts.

Measurement and economics. A substantial empirical literature studies robots at scale: classifier-based analysis of robots usage [37], exclusion as a guidance pro-

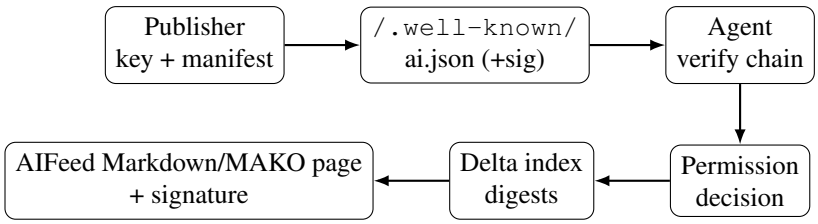

Figure 1: High-level flow: signed manifest and DNS anchor establish attributable permissions; the delta index selects what an agent needs; page documents carry their own signatures bound to the page URL.

tocol [29], robots-based gatekeeping [51], the efficacy of creator protection against AI crawlers (IMC 2025) [41], bot protection for small organizations [30], agent compliance with in-band governance signals [43], and pay-per-crawl pricing [12]. Industry reports supply the traffic asymmetry that motivates this work [21, 23, 48, 52], and a research center documents how AI summaries affect users [46]. On the regulatory side we note the EU AI Act and a 2026 EU feasibility study on a TDM opt-out registry [24, 25], with Indonesia's personal-data protection law as a national example [49]; our legal remarks are open questions, not legal advice.

3 Design

3.1 Trust chain

AIFeed composes four layers, none sufficient alone: TLS (transport identity), domain match (the manifest's `identity.domain` must equal the serving host), an Ed25519 signature over the JCS canonical form of the manifest with domain separation (`aifeed.v0.2\n`), and a DNS TXT anchor (`_aifeed`) binding the public key to the domain. The manifest lives at a well-known URI [45] and is parsed under strict JSON and I-JSON rules [16, 17] with RFC 3339 timestamps [34], JCS canonicalization [50], and BCP 14 key words [15, 38]. Figure 1 summarizes the flow. Clients persist the first accepted key per origin to detect key replacement; the transparency log (RFC 9162 pattern [36]) and key pinning mitigate, but do not eliminate, the trust-on-first-use gap discussed in Section 7.

3.2 Permissions and revocability

The manifest declares a default permission and per-usage values (search, retrieval, input, training, quote, summarize, reproduce, translate, modify, embed, commercial use), attribution policy, and crawl limits. Revocation is served by a registry as

multi-signed documents with due-process statuses (`active`, `under_review`, `suspended`) and bounded staleness: clients treat registry unavailability of more than 168 hours as `UNVERIFIED`. Trust levels are explicit—`VERIFIED`, `UNVERIFIED`, `SUSPENDED`—and high-risk uses require a verified manifest with a DNS anchor.

3.3 Content profiles and per-page binding

AIFeed v0.2 binds policy to individual pages. A page profile carries an `aifeed` block (usage overrides, attribution, limits, license references, and asset links) that is *restrict-only* by default: a page can tighten but never loosen the origin policy. Two profiles are supported over the same signed body bytes:

- AIFeed Markdown v1.0, the native profile [11]: media type `text/aifeed+markdown` (media type procedures [28], markdown media type [39]), extension `.aifeed.md`, marker `aimd: "1.0"`, first-class policy block, optional translation links (`alternates`), and a token budget chosen by the publisher (reference default 4,000).
- MAKO 1.0 [31], the compatibility profile (`text/mako+markdown`); dual-stack origins serve identical bytes under both media types, each with its own signature context (`aimd` or `mako`), which prevents cross-format replay.

A signature container binds the signed bytes to the canonical page URL (`UTF-8("aifeed.aimd.v1\n" || url || LF || raw bytes`); sidecar or inline delivery is permitted. Converters emit referenced media and downloadable files as links (`aifeed.assets`) so agents decide what to fetch, and the safe YAML subset used for frontmatter rejects anchors, aliases, tags, flow collections, and oversized scalars.

3.4 Delta consumption

The delta index (`/.well-known/aifeed-index.json`, signed with context `aimd-index`) lists pages with `sha-256` digests, ETags, token counts, a site resume, and triage fields (title, summary, tags, language, related). Clients diff digests against stored state and fetch only changed pages; unchanged pages cost zero bytes (conditional requests per HTTP semantics [26, 27], content digests per RFC 9530 [47], and index discovery via web linking [44]). Index entries are treated as untrusted claims: digests are verified against fetched bytes before use.

4 Threat Model

We reuse the attack taxonomy evaluated by the conformance suite: manifests hosted on foreign domains, content edited after signing, key replacement, origin or DNS compromise, network modification, stale CDN copies, replays, transport corruption, and local reformatting. The suite also covers content-profile attacks: tampered documents, digest mismatch, cross-URL replay, cross-format/context replay, stripped signatures with manifest policy `signature: required`, fail-open permission overrides, and YAML parser abuses (anchors, aliases, tags, duplicate keys, depth/size bombs). Two properties are explicitly *not* claimed: first-contact origin+DNS compromise is undetectable without out-of-band state; and signatures attest provenance, not the fidelity of a markdown derivative to the HTML rendering. Enforcement is required to make declarations effective: the 1.9 billion bypass events [52] show that unenforced preferences are advisory.

5 Implementation

The reference implementation is zero-dependency [8]. It includes: a strict JSON parser (duplicate keys, NFC, integer bounds) and a safe YAML-subset frontmatter parser; JCS canonicalization and Ed25519 signing/verification; a CLI (`keygen`, `sign`, `validate`, `bundle`, `aimd|mako generate|sign|verify|index|fetch`, `site build`); an npm SDK (`@aifeed/verify`) with manifest, document, index, and triage-selection APIs; an independent Python verifier; a WordPress publisher plugin serving dual-stack content with signed manifests, indices, assets, triage, and `llms.txt`; a static-site builder; and server adapters for eight platforms. Conformance is verified by 25 manifest, 39 MAKO, and 11 AIFeed Markdown vectors executed by both JavaScript and Python implementations, plus PHP differential fixtures and a WordPress end-to-end test. The specifications are published as AIFeed v0.1 [9], v0.2 [10], and AIFeed Markdown v1.0 [11].

6 Evaluation

All numbers below are reproducible from committed artifacts (`benchmarks/*.json`, `npm run bench:mako`, `npm run bench:enforcement`); seeds are fixed, and the environment is a single machine with loopback networking. The corpus is synthetic and deliberately contains navigation, ads, comments, and scripts to model news pages.

Table 1: Publisher-side and AI-side savings versus the no-enforcement baseline (S0); single origin, 18-page site, real HTTP on loopback. "bytes" are response bodies.

	S0	S1	S2	S3
Origin requests	63	45	42	30
Origin bytes	177,537	104,084	95,580	79,555
Origin CPU (ms)	46.6	31.1	28.6	20.4
Peak origin concurrency	17	4	2	2
403 blocks	0	18	18	18
429 limits	0	0	11	11
Client bytes received	177,537	104,570	96,198	80,173
Human p95 latency (ms)	20.0	23.2	19.9	19.3
Publisher byte saving	–	41.4%	46.2%	**55.2%**
Publisher CPU saving	–	33.4%	38.8%	**56.2%**
Peak-concurrency reduction	–	76.5%	88.2%	**88.2%**
AI byte saving (all clients)	–	41.1%	45.8%	**54.8%**
AI byte saving (compliant)	–	42.5%	42.5%	**72.9%**

6.1 Content efficiency

On a 60-page corpus, faithful conversion to the markdown profiles reduces transferred bytes from 1,205,292 to 375,630 (including signature containers), a reduction of **68.83%**; the token estimate (bytes/4) falls by 68.8%. Signing costs 0.25 ms per page; per-page verification costs 0.34–0.70 ms. Delta consumption with 10% of pages changed transfers 51,408 bytes, **95.73%** below the HTML crawl (the model assumes 304 responses for unchanged pages). The MAKO specification claims up to 94% token reduction; that figure presumes semantic summarization by the publisher, which our faithful converter does not perform automatically, and we therefore report the smaller measured conversion value rather than the vendor claim.

6.2 Enforcement harness

A loopback origin behind a policy-enforcement edge (training bots denied with 403; non-compliant AI profiles rate-limited with 429 + Retry-After; humans and compliant clients exempt) serves four deterministic client profiles across four scenarios (Table 1). Scenario S3 combines training denial, rate limiting, and delta consumption.

A 100-tenant run reduces origin bytes from 962,373 to 451,038 in the enforced scenario; per-1,000-tenant projections are included in the artifact but are labeled

as linear extrapolations. Without enforcement (S0) savings are zero by construction, and the bypass literature [52] indicates that this baseline is realistic for non-compliant crawlers.

6.3 Correctness and robustness

Across the conformance suites, signature verification failed zero times, cross-context and cross-URL replays were rejected, and tampered documents were detected by digest mismatch. Frontmatter and container parsers withstood 90,000+ fuzz executions without invariant violations; the safe YAML subset rejects anchor, alias, tag, flow, duplicate, and oversized inputs. The WordPress end-to-end test verifies negotiation, inline signatures, index signatures, assets, triage fields, and `llms.txt` on a real installation.

7 Discussion

Adoption is two-sided and enforcement-dependent. A signed declaration has value only if agents consume and respect it, or if an enforcement point acts on it. We therefore position AIFeed as verifiable input to CDNs, WAFs, and hosting platforms (adapter configurations are provided), not as a replacement for enforcement.

Compatibility profile dependency. The MAKO profile depends on a third-party draft specification; we pin a revision and provide a native profile (AIFeed Markdown) so that the protocol does not depend on a single external document. `terms.txt` and `ai.txt` propose adjacent mechanisms; convergence or divergence will be observable over time.

Trust-on-first-use. Combined origin and DNS compromise on first contact is not cryptographically detectable; mitigations (pinning, anti-rollback, transparency log) require client state or ecosystem participation.

Evaluation scope. Measurements are single-machine, loopback, and synthetic; percentage savings depend on boilerplate ratios and on real crawler behavior. No live 30-day pilot has been completed at preprint time; a pilot kit is published so that independent operators can reproduce and extend the measurement. Human latency effects are bounded by the low-priority policy in our harness but not validated in production.

Legal questions remain open. Machine-readable reservations of rights interact with regimes such as EU TDM opt-out requirements, whose registry feasibility is under study [25]; we flag these as legal questions requiring counsel, consistent with the evidence labels used in the specification.

Conflicts of interest. The authors are the protocol's designers; mitigations are

open specifications, open conformance vectors, committed benchmark artifacts, and reproduction commands. External cryptographic review is planned and not yet complete.

8 Conclusion and Future Work

AIFeed demonstrates that publisher-side, cryptographically attributable content permissions can be specified, implemented without dependencies, and evaluated with modest but reproducible gains: 55–56% publisher-side byte and CPU savings under enforcement, 55–73% AI-side byte savings, and 95.7% savings in steady-state delta consumption, with verification costs under a millisecond per page. The remaining work is not primarily technical: independent cryptographic review, a live pilot, upstream standardization of the trust layer (`EXTENSION.md` for MAKO; an IETF Internet-Draft for the protocol core), and multi-stakeholder governance of the revocation registry. We release all specifications, vectors, and harnesses so that these claims can be challenged, reproduced, and improved.

Ethics and artifact availability

All measurements use synthetic pages and loopback networking; no personal data is processed. The specifications, conformance vectors, fuzzers, benchmark harnesses, and raw results are released under MIT (code), CC BY 4.0 (specification text), and CC0 (vectors). Reproduction commands are listed in the repository README; the public repository URL is recorded in the submission checklist.

References

[1] Associating AI usage preferences with content in HTTP. Internet-Draft draft-ietf-aipref-attach-05, August 2026. Revision 05, 2026-08-19.

[2] A vocabulary for expressing AI usage preferences. Internet-Draft draft-ietf-aipref-vocab-08, September 2026. Revision 08, 2026-09-14.

[3] HTTP message signatures for automated traffic architecture. Internet-Draft draft-meunier-web-bot-auth-architecture-05, March 2026. Revision 05, 2026-03-02.

[4] HTTP message signatures for automated traffic. Internet-Draft draft-ietf-webbotauth-httpsig-protocol-00, September 2026. Revision 00, 2026-09-01; Web Bot Auth working group.

[5] The /llms.txt file, v2. https://llmstxt.org/, 2026. Accessed 2026-09-15.

[6] RSL: Really simple licensing (rsl 1.0). https://rslstandard.org/, 2026. Accessed 2026-09-15.

[7] W3C TDM reservation protocol community group. https://www.w3.org/community/tdmrep/, 2026. Accessed 2026-09-15.

[8] AIFeed Protocol Contributors. AIFeed protocol reference implementation, conformance vectors, and benchmarks, 2026. Public repository URL to be confirmed before submission; see CHECKLIST.md.

[9] AIFeed Protocol Contributors. AIFeed v0.1: Signed declarations for AI-Web content, 2026. spec/en/aifeed-v0.1.md.

[10] AIFeed Protocol Contributors. AIFeed v0.2: MAKO trust profile (extension to v0.1), 2026. spec/en/aifeed-v0.2.md.

[11] AIFeed Protocol Contributors. AIMD v1.0: AIFeed markdown (native content profile), 2026. spec/en/aifeed-aimd-v1.md; media type text/aifeed+markdown.

[12] Richard Archer, Soheil Ghili, and Nima Haghpanah. Pay-per-crawl pricing for AI: The LM-Tree agent, April 2026.

[13] asturwebs. agents.txt: Open standard for AI agent discovery. https://github.com/asturwebs/agents-txt, May 2026. GitHub repository created 2026-05-12, last push 2026-08-29, MIT; inspected 2026-09-15: identity/terms/endpoints file at /agents.txt, not signed, no revocation.

[14] Annabelle Backman, Justin Richer, and Manu Sporny. HTTP message signatures. RFC 9421, February 2023.

[15] Scott Bradner. Key words for use in RFCs to indicate requirement levels. RFC 2119, BCP 14, March 1997.

[16] Tim Bray. The I-JSON message format. RFC 7493, March 2015.

[17] Tim Bray. The JavaScript object notation (JSON) data interchange format. RFC 8259, December 2017.

[18] Rajarshi Chowdhury. terms.txt: A consent and compensation protocol for agentic web access, September 2026. Submitted to IEEE Internet Computing.

[19] Cloudflare. To build a better internet in the age of AI, we need responsible AI bot principles. https://blog.cloudflare.com/building-a-better-internet-with-responsible-ai-bot-principles/, September 2025. Industry report, 2025-09-24; accessed 2026-09-15.

[20] Cloudflare. Giving users choice with Cloudflare's new content signals policy. https://blog.cloudflare.com/content-signals-policy/, September 2025. Industry report; accessed 2026-09-15.

[21] Cloudflare. The crawl before the fall of referrals: understanding AI's impact on content providers. https://blog.cloudflare.com/ai-search-crawl-refer-ratio-on-radar/, July 2025. Industry report, 2025-07-01; accessed 2026-09-15.

[22] Cloudflare. Introducing pay per crawl: enabling content owners to charge AI crawlers for access. https://blog.cloudflare.com/introducing-pay-per-crawl/, July 2025. Product announcement, 2025-07-01; accessed 2026-09-15.

[23] Cloudflare. The 2025 Cloudflare Radar year in review. https://blog.cloudflare.com/radar-2025-year-in-review/, 2025. Industry report; accessed 2026-09-15.

[24] European Commission. AI act: Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai, 2026. Accessed 2026-09-15.

[25] European Commission. New feasibility study for introducing an EU-level registry of text and data mining opt-out. https://digital-strategy.ec.europa.eu/en/library/new-feasibility-study-introducing-eu-level-registry-text-and-data-mini July 2026. Published 2026-07-13; accessed 2026-09-15.

[26] Roy T. Fielding and Julian Reschke. Hypertext Transfer Protocol (HTTP/1.1): Semantics and content. RFC 7231, June 2014.

[27] Roy T. Fielding, Mark Nottingham, and Julian Reschke. HTTP semantics. RFC 9110, STD 97, June 2022.

[28] Ned Freed and John Klensin. Media type specifications and registration procedures. RFC 6838, BCP 13, January 2013.

[29] Dajie Ge and Zhijun Ding. Robots exclusion and guidance protocol. *Tsinghua Science and Technology*, 2016. doi: 10.1109/TST.2016.7787007.

[30] Rama Carl Hoetzlein. Protecting small organizations from AI bots with logrip: Hierarchical IP hashing, August 2025.

[31] Juan Isidoro. MAKO specification v0.1.0 (draft): Markdown agent knowledge optimization. `https://github.com/juanisidoro/mako-spec`, February 2026. Repository commit de7c0d5914751aa52d75ab10714e315bb368196c, 2026-02-18, Apache-2.0.

[32] D. Jolovic. CrawlWall: Block AI crawlers at the edge with Caddy. `https://github.com/jolovicdev/crawlwall`, May 2026. GitHub repository created 2026-05-07, last push 2026-08-15, MIT; inspected 2026-09-15: edge enforcement, bot verification, signed audit ledger and receipts; experimental, no publisher-signed manifest or content profiles.

[33] Simon Josefsson and Ilari Liusvaara. Edwards-curve digital signature algorithm (EdDSA). RFC 8032, January 2017.

[34] Graham Klyne and Chris Newman. Date and time on the internet: Timestamps. RFC 3339, July 2002.

[35] Martijn Koster and Gary Illyes. Robots exclusion protocol. RFC 9309, September 2022.

[36] Ben Laurie, Eran Messeri, and Rob Stradling. Certificate transparency version 2.0. RFC 9162, December 2021.

[37] Junsup Lee, Sungdeok Cha, Dongkun Lee, and Hyungkyu Lee. Classification of web robots: An empirical study based on over one billion requests. *Computers & Security*, 2009. doi: 10.1016/j.cose.2009.05.004.

[38] Barry Leiba. Ambiguity of uppercase vs lowercase in RFC 2119 key words. RFC 8174, BCP 14, May 2017.

[39] Sean Leonard. The text/markdown media type. RFC 7763, March 2016.

[40] Yuekang Li, Wei Song, Bangshuo Zhu, Dong Gong, Yi Liu, Gelei Deng, Chunyang Chen, Lei Ma, Jun Sun, Toby Walsh, and Jingling Xue. ai.txt: A domain-specific language for guiding AI interactions with the internet, May 2025.

[41] Enze Liu, Elisa Luo, Shawn Shan, Geoffrey M. Voelker, Ben Y. Zhao, and Stefan Savage. Somesite i used to crawl: Awareness, agency and efficacy in protecting content creators from AI crawlers. In *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2025. arXiv:2411.15091; accepted at IMC 2025.

[42] Maango. ai-policy.json: Machine-readable AI agent permissions for websites. https://github.com/maango-io/ai-policy.json, April 2026. GitHub repository created 2026-04-18, CC0-1.0; inspected 2026-09-15: consolidated permissions JSON at /.well-known/ai-policy.json, no signature, anchoring, or revocation.

[43] Thamilvendhan Munirathinam. Will the agent recuse, and will it stop? measuring LLM-agent compliance with in-band governance signals, June 2026.

[44] Mark Nottingham. Web linking. RFC 8288, October 2017.

[45] Mark Nottingham. Well-known uniform resource identifiers (URIs). RFC 8615, May 2019.

[46] Pew Research Center. Google users are less likely to click on links when an AI summary appears in the results. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appe July 2025. 2025-07-22; accessed 2026-09-15.

[47] Roberto Polli and Lucas Pardue. Digest fields. RFC 9530, February 2024.

[48] Matthew Prince. Content independence day: no AI crawl without compensation. https://blog.cloudflare.com/content-independence-day-no-ai-crawl-without-compensation/, July 2025. Vendor analysis, 2025-07-01; accessed 2026-09-15.

[49] Republic of Indonesia. Undang-undang nomor 27 tahun 2022 tentang pelindungan data pribadi. https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022, 2022. In force; accessed via peraturan.bpk.go.id 2026-09-15.

[50] Anders Rundgren and Bret Jordan. JSON canonicalization scheme (JCS). RFC 8785, June 2020.

[51] Nicolas Steinacker-Olsztyn, Devashish Gosain, and Ha Dao. Is misinformation more open? a study of robots.txt gatekeeping on the web, October 2025.

[52] TollBit. State of the Bots, H1 2026 (homepage figures). https://www.tollbit.com/, 2026. Industry report; scrapes 22B+ of 987B+ visits, robots.txt bypasses 1.9B+, paywall routings 2.6B+; accessed 2026-09-15.