The Binary Switch: How Websites Govern Autonomous Agent Access

Three measurements of published policy, enforced policy, and defended surfaces, and why operators withhold a vocabulary they already possess

Nicholas, N
Experimental Open Works
xowx.org/contact.html

22 July 2026

Abstract

Autonomous agents now attempt tasks on the web that have consequences: booking, applying, paying, cancelling. We report three web measurements of how sites govern that traffic. First, across the top 500 domains by backlink rank (453 responded) and a hand-picked sample of 40 transactional sites (37 responded), *not one site* publishes any policy on any side-effecting action. The result is not a majority tendency but a census: 0 of 453 and 0 of 37, with all five side-effecting action classes unspecified on 100% of sites. Second, probing five identities per site, we find published policy and enforced policy are close to unrelated. On transactional sites where a browser is served, robots.txt permits an AI identity that the server then refuses in 19.3% of (site, identity) pairs, while robots.txt denials were enforced in 0.0% of pairs; 45% of transactional sites refuse an ordinary browser string outright, before any policy is consulted. Third, classifying 373 robots.txt files into a risk taxonomy shows that operators do possess a granular vocabulary and use it precisely: transactional sites fence inventory and cart state at 4.5× the rate of the web at large, and account, pricing and API surfaces well above it. For AI crawlers the same operators discard that vocabulary: of 136 sites naming an AI crawler, 116 give a blanket `Disallow: /` and only 20 give targeted rules. We argue the barrier to graded agent access is therefore not an absent vocabulary but absent enforcement and absent attribution, and we give a theoretical account of why graded access is nonetheless achievable: the risks operators actually price are defeated by counting against an accountable principal rather than by inferring intent, so the correct unit of policy is the principal, not the request. We report two measurement defects found and corrected mid-study, one analysis error, and an undiagnosed residual sensitivity, and we bound our claims accordingly.

1 Introduction

Software that acts on a person's behalf is no longer hypothetical. An assistant asked to book a hotel room, reschedule a delivery, or cancel a subscription issues ordinary HTTP requests to ordinary websites, and those requests change state in the world. The web's governance vocabulary for automated traffic, however, was designed for a different problem: keeping indexers out of expensive or irrelevant directories. The Robots Exclusion Protocol[1] expresses one predicate, may you fetch this path, and nothing else.

A visible standards effort has grown around this gap. There are now at least seven candidate mechanisms beyond robots.txt by which a site could say something to an agent: a native policy document, Really Simple Licensing (RSL), two incompatible specifications that both claim the path `/agents.txt`, `/agent-manifest.txt`, `/ai.txt` and `/llms.txt`. The prevailing reading of the situation is that agent governance is blocked on a standards problem: the vocabulary is fragmented, so operators cannot express nuance, so they fall back to a binary allow or deny.

We tested that reading and it does not survive contact with the deployed web. This paper reports what we found instead. We measure three things that are usually discussed and rarely counted: what

sites publish, what sites enforce, and what sites protect. The three answers are, respectively, nothing, something unrelated to what they publish, and exactly the risks their business model implies.

The last of these is the one that reframes the problem. Operators are not mute. Confronted with conventional crawlers they write detailed, category-specific rules aimed with precision at cart and checkout state, pricing and search surfaces, account pages and machine APIs. Confronted with an AI crawler the same operators write one line. The vocabulary was available the whole time; a robots.txt group can already say "this crawler may read the catalogue but not the checkout" in two lines, and 116 sites in our sample declined to. We argue that this is a rational response to a system in which no rule binds anything and no violator can be named, and we show, from our own enforcement measurement, that both of those conditions hold.

Contributions.

1. A census of published agent policy over 490 responding domains across two samples, showing that coverage of side-effecting actions is not merely rare but zero, and that the specification fragmentation which motivates much of the current standards work has essentially no deployed instances (Section 5.1).
2. A paired measurement of declared against enforced policy over five identities per site, quantifying the disconnection between robots.txt and the layer that actually decides (Section 5.2).
3. A risk-taxonomy classification of 373 robots.txt files showing that granular policy already exists, is aimed at the predicted risks, and is withheld from AI crawlers specifically (Section 5.3).
4. A theoretical argument that the intent-unobservability objection to graded agent access is misframed, and that the correct unit of policy is the accountable principal rather than the request (Section 6).
5. A candid account of two instrument defects and one analysis error found during the study, and of the residual limitations we did not resolve (Section 8).

2 Problem Formulation

2.1 Actions and dispositions

We model a site's agent policy as a partial function from action classes to dispositions. The action classes, ordered by increasing consequence, are: **read** (fetch pages), **use** (train on or quote content), **write** (create or change things), **communicate** (message a human), **commit** (book, apply or sign), **transact** (spend money) and **destroy** (delete or cancel). The first two are non-side-effecting; the remaining five are side-effecting, and are precisely the classes that arise when an agent is asked to do something rather than to learn something.

Dispositions are **allow**, **deny**, **confirm** (an agent may start this but a human must finish it), **escalate** (needs an account, a contract or a payment) and **unspecified**. A site is *silent* on an action class when every source it publishes leaves that class unspecified. Silence is not neutrality: an agent facing silence must guess, and the consequences of guessing wrong are asymmetric across the seven classes.

2.2 Dial and switch

We call a policy a **dial** when it discriminates among paths, actions or rates, and a **switch** when it is a single global allow or deny for a named party. A robots.txt group containing twelve scoped `Disallow` lines is a dial. A group containing `Disallow: /` is a switch. The distinction is observable in the file itself and requires no inference about the operator's reasoning, which is what makes it measurable.

2.3 The intent objection

The standard argument against graded agent access is that grading requires knowing why a request was made, and motive is not carried in the request. Four adversaries illustrate the point. An agent booking a

room, an agent extracting a competitor's pricing table, an agent walking a checkout flow in order to clone its interface, and an agent holding every middle seat on a flight so that only aisle and window seats remain purchasable all emit request sequences that are, at the protocol level, indistinguishable. If a site cannot tell them apart, the argument goes, its only safe policies are open and closed.

We return to this objection in Section 6, and argue that it misidentifies the quantity that has to be observed.

3 Method

3.1 Instrument

All three experiments use a single open-source measurement tool written in Rust, with three subcommands: `survey` (what a site publishes), `probe` (how a site treats different identities) and `defend` (what a site's robots.txt protects). The tool shares one HTTP client across an entire run, caps every response body, and uses a caching DNS resolver. Section 8 explains why each of those three properties matters more than it should.

3.2 Discovery and source precedence

For each domain the surveyor issues seven concurrent GET requests, to `/.well-known/agent-policy.json`, `/agents.txt`, `/agent-manifest.txt`, `/ai.txt`, `/llms.txt`, `/robots.txt` and `/`. The root document is fetched to discover a declared RSL licence. Sources are normalised into the action-class model of Section 2 and merged by precedence, native policy above RSL, RSL above either `agents.txt` dialect, those above `agent-manifest.txt`, above `ai.txt`, above `llms.txt`, above robots.txt, with a more specific source overriding a less specific one on the same action class.

Two discovery details bear on the results. A site that answers a probe with 401, 403 or 429 is recorded as having refused that probe, and is distinguished from a site that never answered at all; percentages throughout are computed over sites that answered. RSL discovery requires an HTML link element carrying `type="application/rsl+xml"`. An ordinary `rel="license"` link is not sufficient, for reasons given in Section 8.

3.3 Dialect sniffing for `/agents.txt`

Two live specifications claim the path `/agents.txt` with incompatible formats. The IETF draft dialect is plaintext, with a mandatory first significant line of the form `*<sha256>`, followed by `<path> ALLOW` or `<path> DISALLOW` directives, and it specifies that any parse failure means the entire site is restricted. The `agents.txt` v2.0 dialect is Markdown with embedded YAML blocks.

The interaction of those two rules is hazardous. A v2.0 file, read by a conformant parser of the IETF draft, is a parse failure, and therefore reads as a site that has locked every door. A checker that applied failure semantics before determining dialect would report a large fraction of adopting sites as fully closed. Our instrument therefore sniffs first and applies semantics second. It classifies a file as IETF if the first significant line matches `*` followed by 64 hex digits, or, failing that, if the body contains bare `ALLOW` or `DISALLOW` directives (which is the IETF shape without its mandatory hash line, and is reported as an invalid hash rather than as a restriction). It classifies a file as v2.0 on the presence of YAML fences or the characteristic Markdown headings. Anything matching neither is **ambiguous**, and ambiguous is treated as absent, never as restricted. Where an IETF hash line is present, the tool recomputes it as SHA-256 over the file with the hash line, comments and blank lines removed and the remaining lines joined with newlines, and reports agreement or disagreement.

We describe this logic in full because a null result depends on it: a claim that zero sites publish `agents.txt` is only as strong as the detector's ability to recognise both dialects.

3.4 robots.txt resolution

Ground truth for "what the site says" is obtained by parsing robots.txt with standard resolution. The most specific matching user-agent group wins over *, then the longest matching path prefix wins, and `Allow` beats `Disallow` on a tie.

3.5 Identity probes

The prober sends one GET per identity per site, with no retries and no attempt to bypass any control, to public homepages only. Five identities are used: an ordinary Chrome browser string as a control, `GPTBot`, `ClaudeBot`, `PerplexityBot`, and a plain unremarkable bot that declares itself a bot without declaring itself an AI. Each non-browser string retains the real product token, so that a rule keyed on that token fires, and appends our own identifier and a contact URL. Outcomes are classified as **served**, **blocked** (401, 403, 429 or 451), **challenged** (a 200 or 503 carrying an interstitial rather than the page), **server error** or **transport error**. Challenge pages are attributed to a vendor by body signature.

3.6 Risk taxonomy

The `defend` subcommand classifies each `Disallow` path into one of six risk categories plus `other`. Classification is by path segment rather than substring: a path is lowercased and split on non-alphanumeric characters, and a category matches when one of its keywords equals a whole segment. This is why `/search` classifies as pricing and search while `/researcher` does not. Categories are tested in a fixed order so that the most specific wins, and the full keyword lists are reproduced in Appendix A. The categories, with the operator fear each is intended to capture, are **inventory / state** (holding or mutating scarce state; the seat-blocking attack lives here), **pricing / search** (price and catalogue extraction), **account / auth** (identity surfaces), **api / machine** (machine-readable surfaces), **content / media** (bulk content, which is the training-data fear) and **admin / internal** (operator surfaces defended for ordinary security reasons).

A site is counted as naming an AI crawler when any of its user-agent groups matches one of sixteen known AI crawler tokens. Such a site is counted as a **switch** when any of those groups carries `Disallow: /`, and as a **dial** when it names an AI crawler with scoped paths and no blanket denial.

3.7 Samples

Two samples are used throughout. The **general sample** is the top 500 domains by Majestic Million backlink rank, with the top 60 of those used as a control in the identity experiment. The **transactional sample** is 40 hand-picked sites whose business is taking bookings and payments: travel, hotels, retail, ticketing, food delivery and car hire. The second sample exists to correct a known bias in the first, which is discussed in Section 8.

4 Experimental Design

Experiment 1: what sites publish. The surveyor was run over both samples at concurrency 6 with a 12 second per-request timeout, writing one JSON record per domain. The reported quantities are the share of responding sites publishing each source type, and the disposition assigned to each of the seven action classes.

Experiment 2: what sites enforce. The prober was run over the 40 transactional sites and the top 60 general domains, five identities each, with a 400 ms gap between requests to the same host and never more than one request in flight per host. For each (site, AI identity) pair we compare the disposition robots.txt declares for that identity's token against the observed outcome. Sites that refuse the browser control are excluded from the comparison, because for those sites a refusal carries no information about policy.

Experiment 3: what sites protect. One robots.txt fetch per site over both samples. Each `Disallow` path in a group matching * is classified by the taxonomy of Section 3.6; separately, each group naming an AI crawler is classified as switch or dial.

5 Results

5.1 What sites publish

Of the top 500 domains, 453 answered and 90 of those (19.9%) refused at least one probe with 401, 403 or 429. Of the 40 transactional sites, 37 answered. Table 1 gives the published sources.

source	top 500 (453 answered)		transactional (37 answered)	
	sites	share	sites	share
robots.txt	345	76.2%	27	73.0%
llms.txt	42	9.3%	6	16.2%
RSL 1.0	1	0.2%	0	0.0%
`agents.txt` (either dialect)	0	0.0%	0	0.0%
`ai.txt`	0	0.0%	0	0.0%
`agent-manifest.txt`	0	0.0%	0	0.0%
native policy document	0	0.0%	0	0.0%
anything beyond robots.txt	43	9.5%	6	16.2%

Table 1: Published policy sources. Percentages are over responding sites.

Table 2 gives the disposition assigned to each action class across the 453 responding general-sample sites.

class	in plain terms	allow	deny	escalate	unspecified	silent
read	read your pages	309	34	0	110	24.3%
use	train on or quote your content	0	102	1	350	77.3%
write	create or change things	0	0	0	453	**100%**
communicate	message a human	0	0	0	453	**100%**
commit	book, apply or sign	0	0	0	453	**100%**
transact	spend money	0	0	0	453	**100%**
destroy	delete or cancel	0	0	0	453	**100%**

Table 2: Action coverage across the 453 responding sites in the general sample.

Three observations follow. First, the census result: **0 of 453 general sites and 0 of 37 transactional sites say anything about any side-effecting action**. This holds for sites whose entire business is taking bookings and payments. Second, the `deny` column for the **use** class is composed entirely of robots.txt groups blocking named AI crawlers; that is the full extent of the deployed web's vocabulary for anything beyond fetching a page. Third, the specification landscape and the deployed web have almost nothing to do with each other. Adoption of `agents.txt` in either dialect is zero, which means the incompatible-dialect collision described in Section 3.3, a genuine hazard on paper, occurs zero times in 490 responding domains. RSL, publicly backed by a large body of publishers, is deployed on one site.

External validation. Our llms.txt figure of 9.3% for the top 500 sits close to an independently published figure of 8.7% for the top 1,000 as of June 2026. The agreement is the best evidence available to us that the instrument measures what it claims to.

5.2 What sites enforce

The first result of the identity experiment is that on transactional sites, policy is frequently never reached. Table 3 gives refusal rates for the browser control.

sample	sites refusing an ordinary browser string	share
transactional (40)	18	**45%**
top 60 general (control)	12	20%

Table 3: The front door. These sites refuse anything that is not a real browser executing JavaScript with a matching TLS fingerprint, before any policy file is consulted.

Restricting to the 22 transactional sites on which the browser control is served, Table 4 gives refusal by identity.

identity	refused	share of 22
GPTBot	4	18%
PerplexityBot	4	18%
ClaudeBot	3	14%
plain unremarkable bot	1	5%

Table 4: Identity-based refusal on the 22 transactional sites where a browser is served. The sites doing the refusing are airbnb.com, ubereats.com, lyft.com, enterprise.com and ticketmaster.com.

The last row carries the interesting information. Declaring oneself an *AI* bot attracts refusal; declaring oneself an unremarkable bot mostly does not. Identity is already being used as a filter, and it is being used for exclusion only.

Table 5 compares what robots.txt declares for an AI identity against what the server does, over (site, identity) pairs for which both a declared disposition and an observed outcome exist.

	transactional (57 pairs)		top 60 control (120 pairs)	
declared, then observed	pairs	share	pairs	share
robots allows, server refuses	11	**19.3%**	1	0.8%
robots allows, server serves	37	64.9%	85	70.8%
robots denies, server refuses	0	**0.0%**	9	7.5%
robots denies, server serves	9	15.8%	25	20.8%

Table 5: Declared against enforced policy, restricted to sites where the browser control is served. The control sample covers the 48 of 60 sites that served the browser.

The third row is the one to read twice. **No transactional site in this sample with a robots.txt denial actually enforced it**, while a fifth of the time the file explicitly welcomed a bot that the server then turned away. The published file and the component that decides are separate systems, maintained by different people, and in the transactional sample their correlation is close to nil.

Table 6 attributes the interstitials we received.

This matters for who owns the problem. A substantial share of the decisions we measured were not made by the site operator at all, but by a third-party edge vendor configured by them.

5.3 What sites protect

Of the top 500 domains, 373 yielded a usable robots.txt, as did 28 of the 40 transactional sites. Table 7 gives the share of sites disallowing at least one path in each risk category for the * group.

challenge vendor signature	responses
Akamai or comparable WAF	33
Cloudflare	13
DataDome	4

Table 6: Challenge interstitials observed across both probe samples, attributed by body signature.

category	transactional (28)	top 500 (373)	ratio
account / auth	**67.9%**	38.3%	1.8×
pricing / search	**64.3%**	50.1%	1.3×
inventory / state	**57.1%**	12.6%	**4.5**×
api / machine	53.6%	31.4%	1.7×
content / media	32.1%	34.3%	0.9×
admin / internal	28.6%	28.7%	1.0×
other	92.9%	77.7%	1.2×

Table 7: Share of sites disallowing at least one path in each category for `User-agent: *`.

Transactional sites defend inventory state at four and a half times the general rate: carts, baskets, checkouts, bookings, reservations, seats and holds. They defend pricing, search and account surfaces substantially more heavily. They do *not* defend content or media any more than anyone else, and they defend administrative surfaces at exactly the general rate. This is not an inference about what operators fear. It is what they wrote down, in their own files, for their own reasons, before anybody asked them.

Table 8 gives total rule counts across the general sample, which show where the writing effort actually goes.

category	total `Disallow` rules, top 500
other	13,562
pricing / search	2,086
account / auth	667
content / media	608
api / machine	352
inventory / state	278
admin / internal	233

Table 8: Total classified `Disallow` rules for the `*` group across 373 sites.

Table 9 gives the central contrast of this experiment.

Of the 136 general-sample sites that name an AI crawler at all, 116 (85%) use a switch and 20 (15%) use a dial. The same operator who writes twelve carefully scoped `Disallow` rules for `*` writes one line for GPTBot. This cannot be attributed to a missing vocabulary. The file format in which the switch is written is the same file format in which the dial is written, on the same server, for the same crawler, and the dial is two lines longer.

6 The Unit of Policy Is the Principal, Not the Request

Section 2.3 stated the objection that motivates the binary switch: an agent booking a room, an agent extracting a pricing table, an agent learning a checkout flow in order to clone it, and an agent holding every middle seat on a flight to force the sale of aisle and window seats emit identical requests, so no site can grade them. We now argue the objection is misframed. It assumes the quantity to be observed is the

	top 500 (373 with robots.txt)		transactional (28)	
treatment of AI crawlers	sites	share	sites	share
names an AI crawler at all	136	36.5%	9	32.1%
blanket `Disallow: /` (a switch)	116	**31.1%**	3	10.7%
targeted path rules (a dial)	20	5.4%	6	21.4%
blanket-blocks every crawler	37	9.9%	not reported	

Table 9: How AI crawlers are treated in robots.txt, against the same operators' treatment of everyone else in Table 7.

motive behind a request. It is not. In every case on that list, the harm is defeated by *counting against an accountable principal*.

Seat blocking. The attack works only because the two sessions holding the seats are anonymous and unlinked. Bind both to one verified end-user and the finding is arithmetic: six held seats, no purchase, one principal. No motive has been inferred, and no request has been individually judged. The site has counted.

Price extraction. Volume and conversion ratio per principal separate forty lookups followed by a booking from ten thousand lookups followed by nothing. The distinction that matters is not what the requester intended but what the requester did, aggregated. That is still counting.

Inventory holds in general. "One active hold per principal" is expressible, countable and enforceable. It is also entirely meaningless without a principal to count against, which is exactly the condition the current web is in.

Flow cloning. Not defensible, and arguably not worth defending. Anyone can screen record a checkout. A policy layer that promises to prevent this is promising something it cannot deliver, and its failure to deliver would discredit the parts that work.

Training on content. Orthogonal to access control. This is the licensing layer, and RSL already provides a vocabulary for it. Conflating the licensing question with the transaction question is one reason the current debate is stuck: they have different remedies, different beneficiaries and different enforcement paths.

The consequence is a change of predicate. The wrong predicate is *may you book*, which requires knowing why the agent is here. The right predicate is *you may hold one seat, as this verified principal, at this rate, with this consequence if you abuse it*, which requires only identity and arithmetic. Arithmetic has always been available. Cryptographic agent identity and scoped delegation, which supply the principal, became available in 2026. This is why the binary switch was, until now, the correct engineering decision and not an oversight: a rule that nothing enforces, applied to a party that cannot be held accountable, has no value, and the cost of writing it is not zero.

Our third experiment supplies the empirical support for this reframing that the argument would otherwise lack. If operators feared an unbounded and unarticulable range of agent behaviours, the categories they fence would be diffuse. They are not. Transactional operators fence inventory state at 4.5× the general rate and content at 0.9×, which is to say they price precisely the risks that per-principal counting defeats, and do not price the one risk that counting cannot address. The risk model implied by our theoretical account is the risk model operators already wrote down.

7 Discussion

Three findings compose into one argument. Sites can express granular policy and routinely do, aiming it at accounts, pricing and inventory (Section 5.3). Nothing enforces what they express: zero percent of robots.txt denials were honoured by the server in the transactional sample, and nineteen percent of permissions were contradicted by it (Section 5.2). And where voluntary compliance cannot be trusted, operators skip nuance entirely: 85% of those naming an AI crawler use a switch, and 45% of transactional sites refuse even a request carrying an ordinary browser string (Sections 5.3 and 5.2).

The missing piece is therefore not a vocabulary. It is enforcement plus accountability. This has a direct consequence for standards work in this area: a new file format, however well designed, changes nothing on its own, because nothing reads a file at the moment of decision. The decision is made in a WAF, frequently one operated by a third party (Table 6), and the WAF has never read the file. The concrete, demonstrable, per-site problem that exists today is not that operators cannot express what they want; it is that many of them are turning away traffic their own published policy welcomes and do not know it.

There is a corresponding difficulty for anyone proposing to fix this. If 45% of transactional sites refuse anything that is not a real browser, at the edge, the addressable problem in the short term is smaller than the framing suggests, and a meaningful share of it belongs to the edge vendors rather than to site operators. Any serious proposal has to state whether it works with those vendors or around them.

We also note a methodological point that generalises beyond this study. The most important number in this paper, the census of zero, is a null result, and null results are exactly the class of finding that a defective instrument produces most readily. Two of the three defects described in the next section would have produced or preserved a false null. We regard the external validation of the llms.txt figure, and the fact that both samples independently return the same census, as the load bearing evidence that this particular null is real.

8 Limitations and Threats to Validity

We state these prominently, and at length, because several of them were discovered by us during the study and would not otherwise be visible to a reader.

1. **A measurement defect that suppressed detection, found and corrected.** The first run reported only 31% of the top 500 as publishing robots.txt, which is implausible on its face. The cause was that the checker constructed a fresh HTTP client per domain, so there was no connection pooling across the seven probes to a single host and no shared DNS path. Sites answering in 0.03 seconds under `curl` were timing out at the full ten seconds. The fix was to share one client across a run, cap response bodies (root documents are frequently megabytes; policy files never are) and use a caching resolver. robots.txt detection rose from 31% to 76.2%. Every figure in this paper is post-fix.

2. **A measurement defect that manufactured findings, found and corrected.** The first corrected run reported seven Wikimedia domains as declaring an RSL licence that could not then be fetched, which would have been a striking result about live misconfiguration. It was false. RSL discovery matched any HTML `rel="license"` link, and Wikipedia publishes an ordinary Creative Commons licence link. RSL requires `type="application/rsl+xml"`. After the fix, and four regression tests, the finding vanished entirely. We state the lesson plainly because it bears on how this whole class of study should be read: a measurement instrument that manufactures findings is worse than no instrument at all, because its entire value rests on being trusted about something nobody else can see.

3. **An analysis error, corrected.** Our first declared-against-enforced figure was 33%. It was wrong, because it counted sites that refuse all traffic from our address as sites contradicting their own published policy. Restricting the comparison to sites where the browser control is served gives 19.3% (Table 5). The higher figure conflated "blocks our vantage point" with "contradicts its own policy".

4. **Residual concurrency sensitivity, undiagnosed.** On one fixed 60-domain slice, 40, 46 and 51 sites answered at concurrency 20, 10 and 6 respectively. All surveys reported here were run at concurrency 6 for that reason, but the underlying cause is not diagnosed, and we do not claim that concurrency 6 is a fixed point rather than merely better. Any figure derived from this class of instrument at high concurrency should be distrusted, including by us.

5. **Two robots.txt denominators that do not agree.** The survey found robots.txt on 345 of 453 responding sites; the defence classification found a usable robots.txt on 373 of 500 attempted sites. These were separate runs with different denominators and different usability criteria, and we report each within its own experiment rather than reconciling them post hoc. Cross-experiment arithmetic on these two figures is not supported.

6. **A single vantage point.** One IP address and one geography. Edge blocking is frequently address-reputation dependent and region dependent, so every refusal rate in Section 5.2 is specific to where we stood.

7. **Homepages only.** The prober fetches public homepages. Deeper transactional paths, checkout and booking flows in particular, are likely to be more heavily defended, so the practical picture for an agent attempting to complete a task is probably worse than what we measured.

8. **Blocking rates are a floor, not a ceiling.** Our user-agent strings retain the real product token, which is what makes token-keyed rules fire, but append our own identifier and a contact URL. A rule matching a full user-agent string exactly, rather than matching the token, will not fire against us. We therefore under-count refusal.

9. **Sample bias in the general sample.** Majestic ranks by backlinks, which over-weights infrastructure domains (tag managers, static asset hosts, CDN and tracker endpoints) that would never publish an agent policy under any circumstances. The transactional sample exists to correct for this. It returns the same census, which is the more meaningful of the two results.

10. **Small transactional samples.** 40 sites, of which 37 answered the survey, 28 had a usable robots.txt, and 22 served the browser control. The 4.5× inventory result is large enough to be convincing as a direction and is not precise as a magnitude. The same caution applies to every transactional percentage in this paper: a single site moves a transactional share by roughly three points.

11. **Reachability is a floor, not a fact.** 47 of 500 general-sample domains did not answer. Some are genuinely unreachable from our vantage point, some are bot protection, and some may be residual measurement error of the kind described above. All percentages are computed over sites that answered, which is the conservative choice for the census but not necessarily for the others.

12. **`other` is the largest taxonomy bucket**, covering 77.7% of sites and 13,562 rules. Most `Disallow` lines are site-specific paths that no general taxonomy will catch. The comparisons in Table 7 are between named categories and are unaffected by this, but we make no claim that the classification is exhaustive, and a different keyword list would move the absolute shares.

13. **robots.txt is a statement of intent and only that.** Section 5.3 measures what sites say they want. Section 5.2 measures what they do. Nothing in Section 5.3 should be read as a claim about enforcement, and Section 5.2 shows why.

9 Ethics

The measurements reported here were designed to be indistinguishable, in cost to the operator, from ordinary background web traffic. We issued one GET per identity per site, with no retries and no attempt to bypass any access control, to public homepages and well-known policy paths only. Requests to the

same host were separated by 400 ms, with never more than one in flight per host. No authentication was attempted, no state was created or modified on any site, and no data beyond publicly served pages and policy files was collected. Every user-agent string carried a contact URL, so that any operator reading their logs could identify the source of the traffic and reach us. We report refusals as findings about the web, not as obstacles to be worked around, and we treat a site's decision to refuse us as a valid answer to the question we were asking.

10 Related Work

Robots exclusion. The Robots Exclusion Protocol was standardised as RFC 9309[1] after twenty-five years of de facto use. Its predicate is fetch permission on a path, and the standard is explicit that compliance is voluntary. Empirical work on robots.txt long predates AI agents: Sun and colleagues examined bias in robots.txt toward particular search engines[2], and later work measured crawler compliance in the wild[3]. Our third experiment applies the same technique to a new question, which is not who is favoured but which risk is priced.

AI crawler policy at scale. Recent longitudinal work documents a rapid tightening of robots.txt and terms of service against AI crawlers across large web corpora[4]. That work measures the *use* class in our taxonomy. Our contribution is orthogonal: we measure the five side-effecting classes that nothing in that literature covers, and find them empty.

Emerging agent policy formats. Candidate mechanisms include `llms.txt`[5], Really Simple Licensing[6], and two incompatible specifications claiming `/agents.txt`[7,8]. Our survey is, as far as we are aware, the first published adoption measurement covering all of them simultaneously, and it finds the deployment of all but `llms.txt` to be at or near zero.

Agent identity and delegation. HTTP Message Signatures[9] supply the cryptographic substrate for an agent to make verifiable claims about itself, and Web Bot Auth[10] applies them to automated traffic specifically. Scoped delegation from a human principal to software has an established vocabulary in OAuth 2.0[11]. Section 6 argues that the combination of these two, verifiable agent identity bound to a delegating principal, is the enabling condition that the graded-access problem was waiting on.

Measurement ethics. Our protocol follows the guidance of the Menlo Report[12] and of Partridge and Allman on ethics in network measurement[13], in particular the principles of minimising burden on the measured party and making the measuring party identifiable.

11 Conclusion

Nothing on the web says what an agent may do. That is a census, not a tendency: zero of 453 general sites and zero of 37 transactional sites express any policy on any side-effecting action. What sites publish about the actions they do cover barely predicts what their servers do, with robots.txt denials enforced in 0.0% of transactional pairs and robots.txt permissions contradicted in 19.3%. And what sites protect is written with precision, aimed at inventory state, pricing and accounts, and withheld from AI crawlers in favour of a blanket denial in 85% of the cases where an AI crawler is named at all.

The barrier to graded agent access is therefore not an absent vocabulary. Operators possess one, use it daily, and aim it at exactly the risks their business model implies. They decline to extend it to AI agents because a rule that nothing enforces, applied to a party that cannot be held accountable, is worth less than the effort of writing it. The two things that would change that calculus are enforcement that binds a written rule to the component that decides, and attribution that makes abuse traceable to a principal who can be rate-limited or cut off. Both are now technically available. Neither is deployed, and until they are, the switch remains the rational setting.

References

[1] M. Koster, G. Illyes, H. Zeller and L. Sassman. Robots Exclusion Protocol. RFC 9309, IETF, 2022.

[2] Y. Sun, Z. Zhuang, I. G. Councill and C. L. Giles. Determining Bias to Search Engines from Robots.txt. *IEEE/WIC/ACM International Conference on Web Intelligence*, 2007.

[3] C. L. Giles, Y. Sun and I. G. Councill. Measuring the Web Crawler Ethics. *WWW*, 2010.

[4] S. Longpre et al. Consent in Crisis: The Rapid Decline of the AI Data Commons. *NeurIPS Datasets and Benchmarks*, 2024.

[5] J. Howard. The `/llms.txt` file. Proposal, 2024.

[6] Really Simple Licensing (RSL) 1.0. Specification, RSL Collective.

[7] S. Srijal. Agents Policy. `draft-srijal-agents-policy-00`, IETF Internet-Draft.

[8] `agents.txt` v2.0. Specification, MIT licence, asturwebs.

[9] A. Backman, J. Richer and M. Sporny. HTTP Message Signatures. RFC 9421, IETF, 2024.

[10] Web Bot Auth. IETF Internet-Draft, HTTP message signatures for automated agents.

[11] D. Hardt. The OAuth 2.0 Authorization Framework. RFC 6749, IETF, 2012.

[12] D. Dittrich and E. Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. U.S. Department of Homeland Security, 2012.

[13] C. Partridge and M. Allman. Ethical Considerations in Network Measurement Papers. *Communications of the ACM*, 2016.

A Risk taxonomy keyword lists

Categories are tested in the order shown, and the first category with a keyword equal to a whole path segment wins. A path not matching any list is classified as `other`.

category	segment keywords
inventory / state	`cart`, `basket`, `bag`, `checkout`, `order`, `book`, `booking`, `reserve`, `reservation`, `seat`, `availability`, `hold`, `payment`, `pay`, `wishlist`, `waitlist`
pricing / search	`search`, `searchresults`, `results`, `browse`, `filter`, `sort`, `price`, `pricing`, `compare`, `deals`, `offers`, `quote`, `fare`, `rates`, `query`, `listing`, `catalog`, `catalogue`
account / auth	`login`, `signin`, `sign-in`, `logout`, `signout`, `register`, `signup`, `sign-up`, `account`, `profile`, `password`, `auth`, `session`, `member`, `myaccount`, `my-account`, `user`
api / machine	`api`, `graphql`, `ajax`, `json`, `rpc`, `rest`, `webhook`, `callback`, `gateway`
content / media	`image`, `images`, `img`, `media`, `photo`, `video`, `pdf`, `download`, `attachment`, `print`, `amp`, `feed`, `rss`, `export`
admin / internal	`admin`, `wp-admin`, `wp-login`, `cgi-bin`, `internal`, `staging`, `test`, `debug`, `console`, `dashboard`, `backend`, `cms`

We note one dead entry, since we reproduce the lists as they ran rather than as they should have been written. Hyphenated keywords such as `sign-in`, `sign-up`, `my-account`, `wp-admin`, `wp-login` and `cgi-bin` can never match, because segmentation splits on non-alphanumeric characters and a segment therefore never contains a hyphen. Their unhyphenated siblings (`signin`, `signup`, `myaccount`, `admin`) are present in the same lists, so the practical effect is confined to a small number of paths. A site is counted as naming an AI crawler when a user-agent group contains any of the following tokens as a substring: `gptbot`, `chatgpt-user`, `oai-searchbot`, `claudebot`, `anthropic-ai`, `ccbot`, `google-extended`, `perplexitybot`, `bytespider`, `meta-externalagent`, `applebot-extended`, `cohere-ai`, `diffbot`, `omgili`, `amazonbot`, `youbot`.

B Discovery probes and probe identities

Discovery probes. Seven concurrent GET requests per domain: `/.well-known/agent-policy.json` (native policy), `/agents.txt` (both dialects), `/agent-manifest.txt`, `/ai.txt`, `/llms.txt`, `/robots.txt`, and `/` (for RSL link discovery).

Identities. Five identities, one GET each per site. The robots.txt token column gives the group each identity is matched against when resolving declared policy.

identity	role	robots.txt token
browser	control: an ordinary desktop Chrome user-agent string	*
gptbot	named AI crawler	`gptbot`
claudebot	named AI crawler	`claudebot`
perplexity	named AI crawler	`perplexitybot`
plain-bot	self-declared bot with no AI affiliation	our own token

Every non-browser string retains the real product token and appends our own identifier together with a contact URL.

C Reproducibility

The measurement tool is open source, written in Rust, and has three relevant subcommands. All results in this paper were produced with concurrency 6 and a 12 second per-request timeout, and each run wrote one JSON record per domain:

```
remit survey domains.txt --concurrency 6 --timeout 12 --out survey.jsonl
remit probe domains.txt --concurrency 6 --timeout 12 --out probe.jsonl
remit defend domains.txt --concurrency 6 --timeout 12 --out defend.jsonl
```

The general sample is the top 500 domains of the Majestic Million by backlink rank as retrieved on the measurement date, and the top 60 of that list forms the probe control. The transactional sample is 40 hand-picked booking, travel, retail, ticketing, food delivery and car hire domains. Two properties of the runtime are load bearing and should be preserved by any reimplementation: one HTTP client shared across an entire run, so that the seven probes to a host pool their connections, and a response body cap, since root documents are frequently megabytes while policy files never are. Both are discussed in Section 8. Raw JSON records, the domain lists and the analysis are available for research on request through `xowx.org/contact.html`.