Where Over-Blocking Comes From: Simulating CIPA School Filtering Policies on 103 Million Categorised Domains

Samo Plibersek
Alpha Quantum
samo@alpha-quantum.com

October 2026

Abstract

The Children's Internet Protection Act (CIPA) requires US schools and libraries that receive E-Rate support to block visual depictions that are obscene, child sexual abuse material, or harmful to minors. In practice, compliance is implemented with category-based web filters, and districts routinely block far more than the statute names. Concern that filters also block legitimate educational and health content is two decades old, but the evidence comes from small samples of hand-picked sites. We measure the question at census scale. Using web-filtering labels for 103,367,510 domains (57 categories, 9.0% of domains multi-labelled), we simulate five stylised policies, from a statutory core that blocks only adult content to a restrictive district policy that also blocks social, gaming, streaming, shopping and community categories, under three rules for handling sites with several labels. Blocking the statutory core removes 0.92% of domains and catches 0.048% of the 16.3 million domains carrying an education, health, government, news, science or civic label. A typical district policy of 17 categories blocks 4.6% of domains and catches 0.30% of protected-label domains. Adding twelve instructional-time categories quadruples the blocked share to 19.1% and multiplies collateral blocking 26-fold, to 1.24 million protected-label domains (7.6%), including 2.2% of .edu, 1.7% of US K-12 and 4.2% of government domains, none of which the statutory core touches. A cost-ordered blocking curve shows that the statutory and circumvention categories are among the cheapest to block, while Podcasts, Forums, Hate / Discrimination, Blogs, Community Sites and AI / LLM Tools carry the highest collateral ratios (25–60% of newly blocked domains carry a protected label). We also find that the apparent prevalence of adult content depends on the popularity measure: it covers 23% of the global top-1,000 traffic group but at most 0.1% of domains in the upper link-authority bands. Over-blocking, in short, is driven far less by what CIPA requires than by what districts add, and policy choices can be evaluated before deployment with data of this kind. We release all aggregate tables.

Keywords: CIPA, web filtering, over-blocking, school internet policy, E-Rate, multi-label classification, internet measurement

1 Introduction

Since 2001, schools and libraries in the United States that receive E-Rate discounts or certain library grants must certify compliance with the Children's Internet Protection Act (United States Congress, 2000; Federal Communications Commission, 2011). The law asks for a "technology protection measure" that blocks or filters visual depictions that are obscene, child sexual abuse material, or, for minors, harmful to minors, and for an Internet Safety Policy adopted after public notice. The Supreme Court upheld the library provisions in Supreme Court of the United States (2003), relying in part on the ability of adults to ask for the filter to be disabled.

The law names three narrow classes of content, but it is implemented with commercial filters that block by *category*. Districts choose which categories to block, and most choose many more than the statute requires: weapons, drugs, gambling, proxies, social networks, games and streaming are common additions. The worry that this blocks legitimate material is as old as the law. Richardson et al. (2002) showed that pornography filters at their most restrictive settings blocked a substantial share of health information sites; Heins et al. (2006) collected a long record of over-blocking; Batch (2014) described the effect on school libraries ten years after CIPA; Peterson et al. (2017) found that the same sites were treated inconsistently across Alabama schools and libraries; and Laird et al. (2023) reported that most teachers and students see filters block content needed for school work. Resnick et al. (2004) showed how much the measured error rate depends on which sites are tested.

What is missing is a census-scale picture. Most studies test hundreds or thousands of selected URLs, which cannot say how much of the web a policy removes, or which categories cause the damage. Domain-classification services have been studied for consistency (Vallina et al., 2020), but not as inputs to policy. This paper takes the opposite route: instead of testing a filter on a sample, we simulate policies on the full label set of a large commercial domain register and count, exactly, what each policy would block.

Contributions.

1. An exact simulation of five stylised CIPA policies on 103.4 million domains, under three rules for multi-label sites (Tables 1 and 3).
2. A measure of collateral blocking: domains removed by a policy that also carry an education, health, government, news, science or civic label, and the share of institutional domains (.edu, US K-12, government, military) each policy removes (Table 2).
3. A cost-ordered blocking curve that ranks the 28 policy categories by how much protected content each one takes with it (Table 5, Figure 3).
4. Evidence that the measured prevalence of sensitive content depends strongly on the popularity measure, traffic rank or link authority (Figure 4).
5. A public release of every aggregate table used here.

2 Background

CIPA's three content classes are defined by statute. "Harmful to minors" (47 U.S.C. §254(h)(7)(G)) is a sexual-content standard: a depiction that appeals to a prurient interest in nudity, sex or excretion, is patently offensive with respect to minors, and lacks serious value for minors. Violence, weapons, drugs, gambling and hate are therefore not CIPA requirements, although they are routinely blocked under district Internet Safety Policies. The law does not require any particular product or category list, and it does not require blocking for adults in libraries.

Category filters assign each domain one or more labels from a fixed taxonomy. Multi-label assignment (Tsoumakas and Katakis, 2007; Zhang and Zhou, 2014) is natural for the web: a newspaper has a health section, a sports retailer sells hunting rifles. Once a site has several labels, the filter needs a rule. The usual rule is to block if *any* label is blocked; the alternatives are to block only if *all* labels are blocked, or if a majority are. Each rule trades over-blocking against under-blocking, and the trade-off has, to our knowledge, not been measured on a population of this size.

3 Data and method

Register. We use the October 2026 release of a commercial domain register maintained by the author's company: 103,367,510 registrable domains, of which 103,367,500 carry at least one web-filtering label. Labels come from a content classifier applied to each site's pages, followed by a label clean-up pass that limits popular sites to the categories that describe their main purpose. The release uses 57 categories (the taxonomy's 58th, Phishing / Suspicious, has no assignments in this release). 9,272,065 domains (8.97%) carry more than one label. The order of labels within a domain carries no meaning, so we never use a "primary" label. Each domain also has an OpenPageRank score (DomCop, 2026), a 0–10 link-authority measure, a global traffic-rank group for the 11.1 million ranked domains, and, for 41.3 million domains, a country attribution.

Domain classes. From the domain name alone we mark .edu, US K-12 (k12.*xx*.us), government (.gov and national government second-level domains), non-US academic (edu.*xx*, ac.*xx*, sch.*xx* and similar) and .mil domains. These classes are institutional by construction, which makes them a useful check: a policy that blocks them is almost certainly over-blocking.

Policies. We simulate five cumulative policies:

P1 Statutory core: Adult. This is the category that corresponds to CIPA's three classes.

P2 Statutory core + circumvention: P1 plus VPN / Proxy / Anonymizers, without which no filter stays effective.

P3 Broad harm set: P2 plus Lingerie / Swimsuits / Intimate Apparel, Violence, Weapons, Hate / Discrimination, Illegal Drugs and Gambling, the set many vendors and districts present as "harmful".

P4 Typical district policy: P3 plus Alcohol, Tobacco, Dating, File Sharing / P2P, Remote Access Tools, URL Shorteners, Parked / For-Sale Domains, Cryptocurrency and Phishing / Suspicious (17 categories; the last has no domains in this release).

P5 Restrictive district policy: P4 plus twelve categories that districts often restrict during instruction: Social Networking, Gaming, Streaming Media, Forums / Message Boards, Community Sites, Shopping, Marketplaces, Podcasts, Sports & Recreation, AI / LLM Tools, Email and Messaging, and Blogs / Personal Sites.

Protected labels and collateral blocking. We call eight categories *protected*: Education/Reference, Health & Medicine, Government & Legal, News/Media, Science, Webinars/E-learning, Religion & Spirituality and Political. 16,250,357 domains carry at least one of them. A blocked domain that also carries a protected label is counted as *collateral*. This is a measure of exposure, not of error: a site can legitimately be both a shop and a health resource. But every such block removes content that a school has a reason to keep open, and the measure is directly comparable across policies.

Computation. A single streaming pass over the register counts domains per distinct combination of label set, OpenPageRank band, traffic-rank group and domain class (about 427,000 combinations). Every figure in this paper is computed exactly from these counts; no sampling is involved.

Table 1: Domains blocked by five stylised policies under the any-label rule, and the protected-label domains they catch. Protected labels: Education/Reference, Health & Medicine, Government & Legal, News/Media, Science, Webinars/E-learning, Religion & Spirituality, Political ($N_{\text{prot}} = 16{,}250{,}357$).

	Policy	Cat.	Blocked	% of all	Protected caught	% of protected
P1	Statutory core	1	948,020	0.92	7,756	0.048
P2	Statutory core + circumvention	2	962,376	0.93	7,937	0.049
P3	Broad harm set	8	3,469,778	3.36	27,808	0.171
P4	Typical district policy	17	4,743,815	4.59	48,046	0.296
P5	Restrictive district policy	29	19,726,461	19.08	1,236,688	7.610

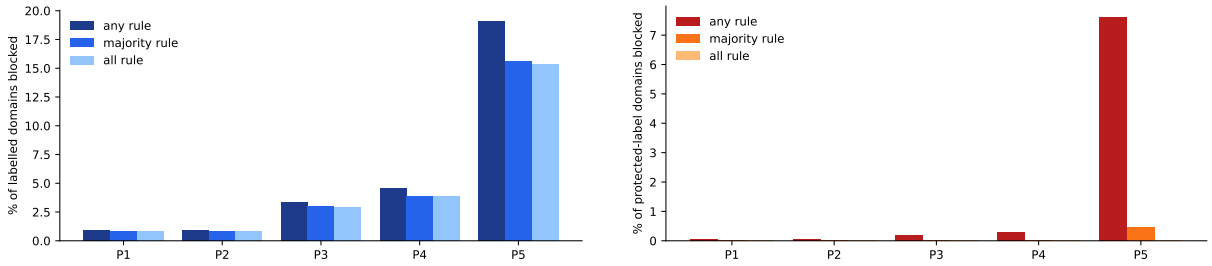


Figure 1: Left: share of labelled domains blocked by each policy under the three rules. Right: share of the 16.3 million protected-label domains that each policy blocks.

4 Results

4.1 How much each policy blocks

Table 1 and Figure 1 give the headline. The statutory core blocks 948,020 domains, 0.92% of the register. Adding circumvention barely changes this (0.93%), because VPN and proxy sites are few (14,771 domains), although each one matters disproportionately. The broad harm set more than triples the blocked share to 3.36%, and the reason is a single category: Gambling, with 2,266,884 domains, is more than twice as large as Adult. The typical district policy blocks 4.59%. The restrictive policy jumps to 19.08%, one domain in five.

4.2 Collateral blocking

Collateral blocking is small, and nearly flat, up to the typical district policy: 7,756 protected-label domains for the statutory core (0.048% of all protected-label domains), 27,808 for the broad harm set (0.17%) and 48,046 for the typical district policy (0.30%). The restrictive policy catches 1,236,688 (7.61%), 26 times the typical policy. The cliff between P4 and P5 is the most important result of this paper: the over-blocking that schools complain about is produced almost entirely by the categories districts add for instructional reasons, not by the categories CIPA requires.

The institutional domains in Table 2 tell the same story without any reliance on the protected-label definition. No .edu or US K-12 domain is blocked by the statutory core, and only nine government domains are. Under the restrictive policy, 148 of 6,596 .edu domains (2.2%), 75 of 4,398 US K-12 domains (1.7%) and 2,802 of 67,058 government domains (4.2%) are blocked. The labels responsible are not the sensitive ones: Sports & Recreation accounts for 63% of the blocked K-12 domains and 49% of the blocked .edu domains (school athletics), Shopping for a further 25% of .edu (campus stores), and Community Sites for 61% of the blocked government domains (town and county portals).

Table 2: Institutional domains blocked under the any-label rule. "Government" covers .gov and national government second-level domains (gov.*xx*, gob.*xx*, gouv.*xx*, govt.*xx*, go.*xx*); "academic (non-US)" covers edu.*xx*, ac.*xx*, sch.*xx*, school.*xx* and k12.*xx*.

Domain class	Domains	P1	P2	P3	P4	P5
.edu	6,596	0 (0.00%)	0 (0.00%)	1 (0.02%)	2 (0.03%)	148 (2.24%)
k12.*xx*.us	4,398	0 (0.00%)	0 (0.00%)	4 (0.09%)	5 (0.11%)	75 (1.71%)
government	67,058	9 (0.01%)	9 (0.01%)	185 (0.28%)	241 (0.36%)	2,802 (4.18%)
academic (non-US)	126,262	95 (0.08%)	98 (0.08%)	361 (0.29%)	550 (0.44%)	2,276 (1.80%)
.mil	87	1 (1.15%)	1 (1.15%)	1 (1.15%)	1 (1.15%)	3 (3.45%)

Table 3: Effect of the blocking rule. *Any*: block if any label is in the policy. *Majority*: block if more than half of the labels are. *All*: block only if every label is. "Missed counts domains that carry at least one policy label but stay open.

	Any		Majority		All	
Policy	Protected caught	Missed	Protected caught	Missed	Protected caught	Missed
P1	7,756	0	0	85,772	0	85,772
P2	7,937	0	1	92,095	0	92,167
P3	27,808	0	208	412,292	0	417,034
P4	48,046	0	1,107	767,524	0	783,488
P5	1,236,688	0	74,791	3,612,050	0	3,907,124

4.3 The blocking rule

Table 3 compares the rules. The *all* rule removes collateral blocking entirely by construction, and the *majority* rule almost entirely, but both leave mixed sites open: 85,772 adult-labelled domains under P1 and 3.9 million policy-labelled domains under P5. For the statutory core these misses are exactly the sites that CIPA requires to be blocked, so the *any* rule is the only defensible choice there. For instructional categories the calculus is different: under the *majority* rule, P5's collateral falls from 1.24 million to 74,791 domains. A filter that applies the strict rule to the statutory core and a lenient rule to discretionary categories would remove most of the collateral damage while keeping the legal floor intact. To our knowledge, no common filter exposes this choice.

4.4 What sensitive categories co-occur with

Table 4 and Figure 2 show why categories differ so much in collateral cost. Adult and Gambling are almost always the only label on their domains (91.0% and 88.0%) and rarely co-occur with protected labels (0.8% and 0.3%). Blocking them takes little else with it. Hate / Discrimination is the opposite: 37.1% of its domains carry a protected label, mostly Political (29.9%), a pattern consistent with advocacy, commentary and history sites that discuss hate rather than promote it. Illegal Drugs co-occurs with Health & Medicine on 12.2% of its domains, the overlap where treatment, prevention and pharmacology sites sit. AI / LLM Tools is a special case: only 3.0% of AI domains carry no other label, and 15.9% are also Education/Reference, so a blanket AI block in a school hits tutoring and learning tools as well as chatbots. Weapons co-occurs with Shopping, Sports and Military, and Lingerie / Swimwear is mostly retail (36.7% Shopping).

4.5 A cost-ordered blocking curve

If a district wanted to block as much as possible while sacrificing as little protected content as possible, in which order should it add categories? We answer this greedily: at each step we

Table 4: What the sensitive categories co-occur with. Single: share of the category's domains with no other label. Protected: share that also carries a protected label. The three most frequent co-labels are shown with the share of the category's domains that carry them.

Category	Domains	Single	Protected	Most frequent co-labels
Adult	948,020	91.0%	0.8%	Dating 3.4%, Arts & Ent. 1.1%, Shopping 0.9%
Lingerie / Swimwear	60,092	52.3%	2.4%	Shopping 36.7%, Adult 12.9%, Health 2.1%
Violence	30,613	92.3%	4.4%	News/Media 1.9%, Arts & Ent. 1.3%, Political 1.1%
Weapons	103,793	77.4%	3.1%	Shopping 9.6%, Sports 6.8%, Military 6.6%
Hate / Discrimination	3,053	53.0%	37.1%	Political 29.9%, Violence 6.3%, Community 4.8%
Illegal Drugs	54,734	72.4%	13.7%	Shopping 12.9%, Health 12.2%, Tobacco 1.7%
Gambling	2,266,884	88.0%	0.2%	Gaming 9.7%, Sports 1.7%, Cryptocurrency 0.2%
VPN / Proxy	14,771	51.7%	1.2%	Computers & Tech. 17.4%, Gambling 11.3%, CDN / Hosting 5.3%
AI / LLM Tools	140,781	3.0%	23.3%	Business 55.5%, Computers & Tech. 47.1%, Web apps 22.2%

add the policy category whose *newly* blocked domains carry the lowest share of protected labels. Table 5 and Figure 3 show the result.

Parked domains, Gambling, URL Shorteners, Email and Messaging, Adult, Alcohol and VPN / Proxy are the cheapest categories: below 1.4% of their newly blocked domains carry a protected label. All CIPA-relevant categories sit in the cheap part of the curve. The curve stays nearly flat to about 6% of the web, then Shopping adds 8.0 million domains at once and the slope turns upward. The expensive tail is dominated by instructional categories: Community Sites (0.26), Blogs (0.28), Forums (0.42) and Podcasts (0.60), with AI / LLM Tools at 0.25 and Hate / Discrimination at 0.35. These are the categories where per-site review, allow-lists or time-based rules pay off most.

4.6 Popularity: two measures, two pictures

Adult content looks very different depending on how popularity is measured (Figure 4). By traffic rank it is a head phenomenon: 151 of the 646 domains in the global top-1,000 group (23.4%) are labelled Adult, 15.4% of the next group (1,001–5,000) and 13.9% of ranks 5,001–10,000, against 0.88% of unranked domains. By link authority it is a tail phenomenon: 0.96% of domains with OpenPageRank 0–1, 0.11% at 5–6 and none above 7. Adult sites attract enormous traffic but few editorial links. The practical consequence is methodological: a filter evaluation built on a link-authority list (as most "top sites" lists derived from crawls are) will underestimate the exposure that matters to students, and one built on a traffic list will make the adult share of the web look 25 times larger than it is. Figure 5 shows the same divergence for the other policies: blocked shares fall with link authority for P1–P3, while P5, driven by social and streaming platforms, rises to 25.5% of domains in band 6 and higher still in the sparsely populated bands above.

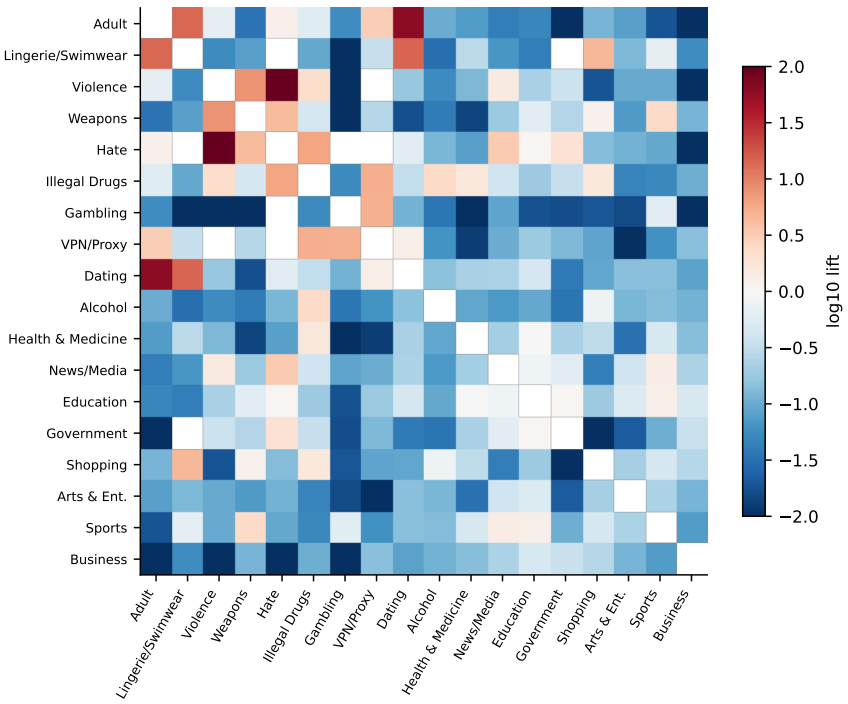


Figure 2: Co-occurrence lift (observed over expected under independence, log10) between the broad harm categories and selected context categories. Red cells co-occur more often than chance.

4.7 Geography and domain suffixes

Country attribution is available for 41.3 million domains. Among countries with at least 100,000 attributed domains, China-attributed domains have the highest Adult share (6.3%), against 0.72% for Russia and 0.44% for the United Kingdom. Suffixes (Table 6) separate even more clearly: .com.cn (9.5%) and .cn (7.0%) lead, followed by inexpensive generic suffixes such as .cc, .top, .vip and .xyz (4.6–5.9%), where the broad harm share reaches 27% (.vip). Country-code suffixes with registration requirements, such as .ie, .com.tr, .co.jp, .no and .com.au, stay below 0.2%. Suffix-level rules are a blunt instrument, but for schools in particular a review of cheap generic suffixes is a defensible complement to category blocking.

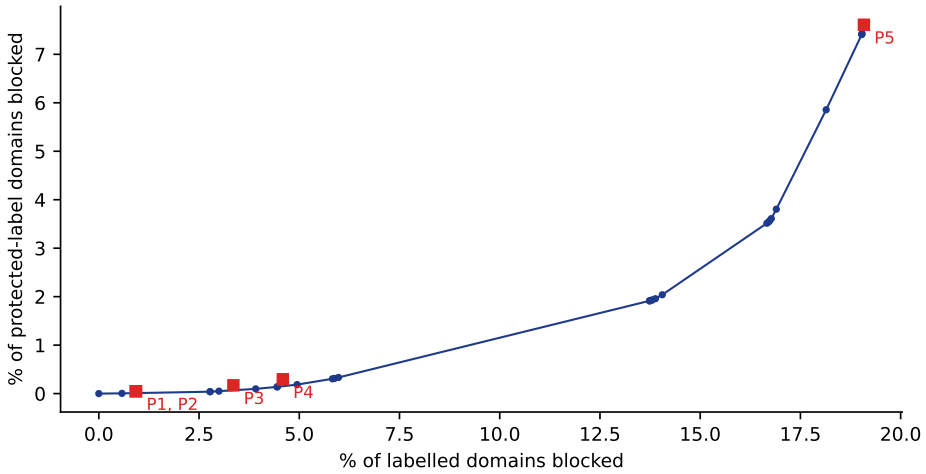


Figure 3: Cost-ordered blocking curve over the 28 policy categories. Each point adds one category. Red squares: the five policies of Table 1 (P1 and P2 overlap).

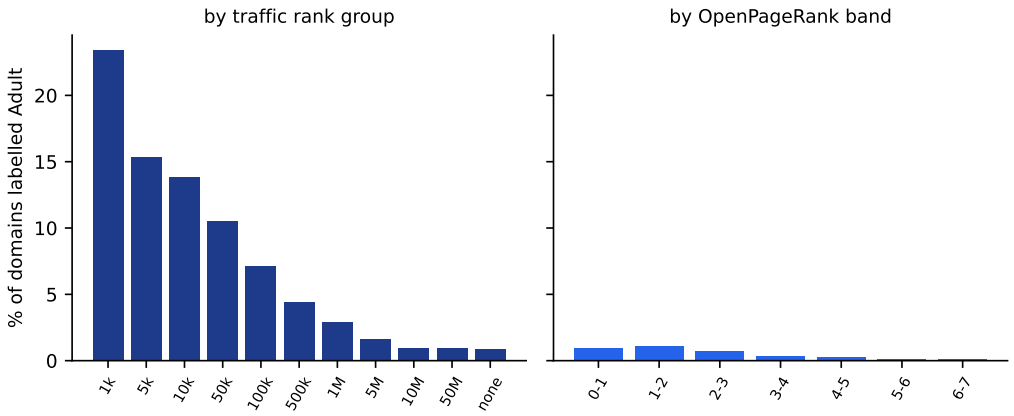


Figure 4: Share of domains labelled Adult by global traffic-rank group (left) and by OpenPageRank band (right; bands with at least 1,000 domains).

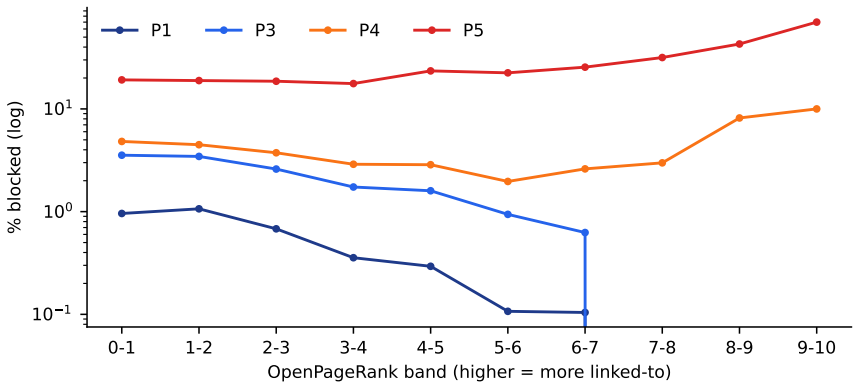


Figure 5: Share of domains blocked under the any-label rule by OpenPageRank band (log scale). Bands above 7 contain fewer than 400 domains each.

Table 5: Cost-ordered blocking curve over the 28 policy categories present in the data. At each step the category whose newly blocked domains carry the lowest share of protected labels is added. Ratio: protected share among the newly blocked domains of that step.

Step	Category added	Tier	Newly blocked	Ratio	Cum. % protected
1	Parked / For-Sale Domains	district	596,529	0.001	0.00
2	Gambling	broad harm	2,266,799	0.002	0.04
3	URL Shorteners	district	10,177	0.006	0.04
4	Email and Messaging	instruct.	224,861	0.008	0.05
5	Adult	core	946,619	0.008	0.10
6	Alcohol	district	548,193	0.012	0.14
7	VPN / Proxy	circumv.	12,403	0.013	0.14
8	Social Networking	instruct.	506,519	0.015	0.19
9	Gaming	instruct.	911,284	0.021	0.30
10	Lingerie / Swimwear	broad harm	52,283	0.026	0.31
11	Weapons	broad harm	103,303	0.031	0.33
12	Shopping	instruct.	8,019,009	0.032	1.91
13	Remote Access Tools	district	2,176	0.032	1.91
14	File Sharing/P2P	district	12,951	0.036	1.92
15	Violence	broad harm	30,036	0.042	1.92
16	Tobacco	district	22,440	0.046	1.93
17	Marketplaces / Classifieds / Auctions	instruct.	85,364	0.050	1.96
18	Streaming Media	instruct.	176,498	0.075	2.04
19	Sports	instruct.	2,697,182	0.089	3.51
20	Illegal Drugs	broad harm	45,909	0.098	3.54
21	Dating	district	22,170	0.149	3.56
22	Cryptocurrency	district	45,643	0.167	3.61
23	AI / LLM Tools	instruct.	129,417	0.245	3.80
24	Community	instruct.	1,284,905	0.259	5.86
25	Blogs	instruct.	916,915	0.275	7.41
26	Hate / Discrimination	broad harm	2,520	0.346	7.41
27	Forums / Message Boards	instruct.	5,474	0.422	7.43
28	Podcasts	instruct.	48,882	0.604	7.61

Table 6: Domain suffixes with at least 100,000 domains: highest and lowest shares labelled Adult, with the share carrying any label of the broad harm set (Adult, Lingerie / Swimwear, Violence, Weapons, Hate, Illegal Drugs, Gambling).

Suffix	Domains	Adult	Harmful	Suffix	Domains	Adult	Harmful
.com.cn	120,718	9.53%	10.3%	.ie	122,551	0.09%	0.6%
.cn	468,680	6.95%	8.3%	.com.tr	152,835	0.09%	0.4%
.cc	108,320	5.87%	16.3%	.ir	144,131	0.11%	0.3%
.top	366,635	5.62%	19.0%	.dev	137,756	0.14%	1.1%
.vip	163,659	4.82%	27.4%	.co.jp	263,279	0.16%	0.3%
.xyz	709,014	4.64%	12.0%	.no	285,349	0.16%	0.8%
.club	112,308	3.95%	21.3%	.tech	113,539	0.17%	1.6%
.pro	200,780	2.50%	10.0%	.co.za	321,905	0.19%	0.8%

5 Discussion

For districts. The statutory floor is cheap. Blocking Adult and VPN / Proxy removes under 1% of the web and almost nothing a school needs. Even the typical 17-category district policy catches under 0.3% of protected-label domains. The damage starts when instructional categories are *blocked* rather than *managed*: Community Sites, Blogs, Sports, Shopping and Podcasts carry the civic, athletic and educational life of schools and towns. Time-based rules, per-grade profiles and allow-lists achieve the instructional goal with far less collateral than a block. The same applies to AI tools, where a blanket block removes tutoring tools with the chatbots.

For E-Rate audits. Auditors check that a technology protection measure exists and that an Internet Safety Policy was adopted; they do not measure over-blocking. The numbers here suggest a simple reporting standard: the share of institutional domains (.edu, K-12, government) a configuration blocks. A CIPA-only configuration blocks none of them.

For vendors and researchers. Filters should expose every label a site carries and let administrators choose the multi-label rule per category. Evaluations of filters should state which popularity measure their test list follows, because traffic and link authority give opposite pictures of the sensitive web.

6 Limitations

Labels are classifier outputs from one commercial register; we do not measure per-category precision here, and errors in rare categories (Hate / Discrimination has 3,053 domains) weigh more than errors in common ones. The analysis is at domain level: a news site's single article or a forum's single thread is not seen. "Protected" is a definition, and collateral is exposure, not proof of error. The five policies are stylised; real configurations differ by vendor and district. Traffic-rank groups cover 11.1 million domains and country attribution 41.3 million, so those breakdowns describe subsets. Finally, all figures are a snapshot of October 2026.

7 Conclusion

Measured over 103 million domains, the content CIPA actually requires schools to block is a small, self-contained part of the web: blocking it, together with circumvention tools, removes under 1% of domains and essentially no educational, health or civic sites. Over-blocking enters with the categories districts add to manage instruction, which multiply collateral blocking more than twentyfold. The policy lesson is to keep the statutory floor strict and to manage everything else with rules that can tell a school's athletics page from a distraction.

Data availability

All aggregate tables behind this paper (category prevalence by popularity band, the full 57×57 co-occurrence matrix, policy simulations by rule, domain class and popularity, the cost-ordered blocking curve, sensitive-category context, and country and suffix shares) are released as an open dataset under CC BY 4.0. No domain-level labels are included.

Declaration on the use of AI

AI-assisted tools were used to help write the analysis code, draft and edit the text and prepare figures. The author reviewed all content and takes full responsibility for it.

Competing interests

The author is affiliated with Alpha Quantum, which maintains the domain register analysed here and operates Web Filtering Database (`https://www.webfilteringdatabase.com`) and CIPA Web Filtering (`https://www.cipawebfiltering.com`), which offer web filtering data and a hosted filtering service for schools and libraries.

References

Kristen R. Batch. Fencing out knowledge: Impacts of the children's internet protection act 10 years later. Policy Brief 5, American Library Association, Office for Information Technology Policy, 2014.

DomCop. Open pagerank. Domain authority scores derived from the Common Crawl link graph, 2026.

Federal Communications Commission. Children's internet protection act certifications. 47 C.F.R. §54.520, 2011.

Marjorie Heins, Christina Cho, and Ariel Feldman. Internet filters: A public policy report. Technical report, Brennan Center for Justice, NYU School of Law, 2006.

Elizabeth Laird, Maddy Dwyer, and Hugh Grant-Chapman. Off task: EdTech threats to student privacy and equity in the age of AI. Technical report, Center for Democracy & Technology, 2023.

Chris Peterson, Shannon M. Oltmann, and Emily J. M. Knox. The inconsistent work of web filters: Mapping information access in Alabama public schools and libraries. *International Journal of Communication*, 11:4583–4609, 2017.

Paul J. Resnick, Derek L. Hansen, and Caroline R. Richardson. Calculating error rates for filtering software. *Communications of the ACM*, 47(9):67–71, 2004. doi: 10.1145/1015864.1015866.

Caroline R. Richardson, Paul J. Resnick, Derek L. Hansen, Holly A. Derry, and Victoria J. Rideout. Does pornography-blocking software block access to health information on the Internet? *JAMA*, 288(22):2887–2894, 2002. doi: 10.1001/jama.288.22.2887.

Supreme Court of the United States. United states v. american library association, inc. 539 U.S. 194, 2003.

Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3):1–13, 2007. doi: 10.4018/jdwm.2007070101.

United States Congress. Children's internet protection act. Pub. L. 106-554, codified at 47 U.S.C. §254(h) and 20 U.S.C. §9134, 2000.

Pelayo Vallina, Victor Le Pochat, Álvaro Feal, Marius Paraschiv, Julien Gamba, Tim Burke, Oliver Hohlfeld, Juan Tapiador, and Narseo Vallina-Rodriguez. Mis-shapes, mistakes, misfits: An analysis of domain classification services. In *Proceedings of the ACM Internet Measurement Conference (IMC '20)*, pages 598–618, 2020. doi: 10.1145/3419394.3423660.

Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014. doi: 10.1109/TKDE.2013.39.