The Commons Captured

The Commons Captured

A Critical History of Wikipedia's Promise

Michael J Bommarito II

First Edition
2026

The Commons Captured:
A Critical History of Wikipedia's Promise

Cover design and typesetting by the author. Printed in the United States of America.

Book Website: `commonscaptured.com`
Author Website: `michaelbommarito.com`

ISBN 979-8-951640-00-0 (paperback)

ISBN 979-8-951640-01-7 (ebook)

Publisher's Cataloging-in-Publication Data
Names: Bommarito, Michael J., II, author.
Title: The commons captured : a critical history of Wikipedia's promise / Michael J Bommarito II.
Description: First edition. | Michigan : 2026. | Includes bibliographical references.
Identifiers: ISBN 979-8-951640-00-0 (paperback) | ISBN 979-8-951640-01-7 (ebook)
Subjects: LCSH:
Wikipedia.
Wikimedia Foundation.
Electronic encyclopedias—History.
User-generated content.
Open source software.
Copyleft.
Artificial intelligence—Social aspects.
Information commons.

Classification: LCC ZA4482 .B66 2026 | DDC 030—dc23

To the volunteers who built the commons,
and those who continue to defend it

We need to launch a movement to develop a universal free encyclopedia, much as the Free Software movement gave us the free software operating system GNU/Linux. The free encyclopedia will provide an alternative to the restricted ones that media corporations will write.

Richard Stallman
"The Free Universal Encyclopedia
and Learning Resource," 1999

Contents

Preface

In 1999, the programmer Richard Stallman imagined an encyclopedia that could not be enclosed.[1] He had spent the better part of two decades watching corporations take freely shared software, modify it, and lock away the improvements.[2] The solution, he argued, was a license that required sharing back. Any work built on the commons would become a permanent part of the commons. Stallman called this copyleft.[3]

Two years later, Wikipedia launched under that principle.[4] Its original license, the GNU Free Documentation License (GFDL), required that derivative works remain free in the same sense as the original.[5] Anyone could use Wikipedia's content. Anyone could improve it.[6] No one could capture it.

Twenty-five years on, that promise has been hollowed out.

Every major large language model trains on Wikipedia.[7,8] OpenAI's GPT-3 drew 3 percent of its training data from the encyclopedia.[7] Meta's LLaMA used 4.5 percent.[8] Researchers have shown that these models reproduce Wikipedia's text, often word for word, and, even when the output is not an exact copy, in unmistakable fragments of the original articles.[9,10,11] Among ChatGPT's ten most-cited sources, Wikipedia alone accounts for 47.9 percent of citations—more than any other site.[12] The encyclopedia that a quarter million volunteers maintain has become the factual backbone of a trillion-dollar industry.[13,14]

The license was supposed to prevent exactly this. Both the original GFDL and the Creative Commons Attribution-ShareAlike license (CC-BY-SA) that replaced it in 2009 require attribution and share-alike.[15] If you build on Wikipedia, you must credit it. If you create a derivative work, that work must carry the same freedoms.[16]

But when a language model ingests Wikipedia and generates text, where is the attribution? Where is the share-alike? On paper, the terms still apply. Whether they bind the companies that train models on Wikipedia is contested—but no one ever tried to enforce them, so the question was never tested.[17]

The Wikimedia Foundation knows this. In March 2024, the team behind the Foundation's paid data service, Wikimedia Enterprise, acknowledged in writing that many AI companies using Wikimedia data might not be complying with either the letter or the spirit of the license.[18]

But instead of enforcing the license, the Foundation kept selling. Through its Enterprise arm, it now counts Google, Amazon, Meta, and Microsoft among its paying clients.[19] The Foundation collects fees while the AI companies capture billions in market value built on volunteer labor.[20] The editors who wrote the articles receive nothing. The readers who might have visited Wikipedia visit ChatGPT instead. Human pageviews to the encyclopedia fell 8 percent year over year in 2025, with the decline steepening as AI search spreads.[21]

A question runs underneath all of this: why did the institution built to defend the commons decline to defend it? Part of the answer is financial. Part is cultural. The people who ran the Foundation moved through the same conferences, credentials, and politics as the executives across the table. The pattern is measurable, and it runs one way. Across eighteen years of federal records, more than 99 percent of Foundation employees' identifiable political giving went to a single party. You can probably guess which.[22]

Yet whether you caricature the staff as donkeys or elephants, the color of the partisanship is not the point. What matters is that a monoculture of any stripe weakens an institution from both directions at once. Inside, it strips out the friction that produces debate, so no one is left to argue the unpopular case. Outside, it makes the institution cheap to capture: per-

suade the one worldview everyone already shares, and you have persuaded everyone.

Had the giving run 99 percent the other way, the exposure would be the same. The homogeneity left the Foundation without an internal opposition, without anyone inclined to treat the AI companies as adversaries rather than partners. When the extractors arrived, it did not see a threat—it saw people like itself.

That was the first capture. It left Wikipedia open to the second.

This is a book about that transformation—how a project founded on copyleft principles became AI infrastructure, and what happens when the institution created to protect a commons begins to profit from its capture.

We trace the arc from Stallman's copyleft idea through Wikipedia's founding, its licensing controversies, its institutional growth, and its current crisis.[1] We examine the 2009 license migration, a legally questionable maneuver that may have broken the guarantees copyleft was supposed to provide.[15] We follow the Wikimedia Foundation's evolution from a shoestring operation to a $180 million institution with an enterprise sales arm.[23,20] We measure the political homogeneity of its staff. And we document the factions now forming over AI and commercialization.[24]

By early 2026, the paradox was visible. Grokipedia, a fork of Wikipedia built in the open, with free access and attribution, seeded its articles with neo-Nazi sources and drew public condemnation from the Foundation.[25,26,27] ChatGPT, which provided no attribution and charged for access, began citing Grokipedia as authoritative—and drew no objection at all.[28] The Foundation condemned Musk's public, compliant fork and kept accommodating the private extraction that broke its license.

[1]This account draws on Wikimedia Foundation documents, Wikipedia's own records (policy pages, deletion logs, arbitration records, talk-page debates), academic research, contemporary news coverage, and web archives, with every factual claim cited. For a fuller discussion of sources and verification, see "A Note on Sources and Methods" in the back matter.

This is not a polemic against Wikipedia. Wikipedia remains one of humanity's great collaborative achievements. But achievements require protection. And somewhere between the volunteer editors and the AI boardrooms, protection failed.

Who benefits from Wikipedia's value? The volunteers who created it? The Foundation that hosts it? Or the corporations extracting from it at scale? The answer is not what Wikipedia's founders intended.

The people who built Wikipedia deserve a clear-eyed account of what they built and what became of it. So do the rest of us who rely on their work.

Michigan
June 2026

Prologue

You can learn a lot about an institution from the rules it creates and the way it chooses to enforce them, or not.

In March 2024, the Wikimedia Foundation admitted in writing that many of the AI companies using Wikipedia might be violating its license.[18] That license carried the promise the whole project rested on: what is built from shared work stays shared. The Foundation enforced nothing, and its commercial arm kept collecting from the industry it suspected.[29] In June 2026, the same institution expelled the man who had co-founded Wikipedia and given it its name.[30] His offense, in the end, was disagreement.

To understand those two choices, it helps to begin with a copy.

On October 27, 2025, Elon Musk's xAI launched a website called Grokipedia. It was a near-complete copy of Wikipedia, produced by running the encyclopedia through Musk's AI and rewriting it to correct what Musk and many others saw as systematic bias in Wikipedia's sources and framing.[31] At launch, though, the AI "balanced" that bias by pulling in sources Wikipedia's own rules excluded, among them InfoWars and neo-Nazi forums.[32]

The motivation deserved a hearing. The worry that Wikipedia had developed a political lean, in its articles and in its rulings on which sources counted as reliable, was shared by many readers and critics.[33] The exe-

cution deserved none. An AI set loose on a reference work, correcting it against the open internet, produced exactly the contamination Wikipedia's standards existed to prevent. It was crude and easy to condemn, and the Foundation condemned it.[27]

But condemnation was all the Foundation offered. The problem was the Foundation's passivity, not Musk himself. Musk had merely done in public what others had been doing in private for years. For four years the Foundation had signed deals with AI companies that trained on Wikipedia without meaningful attribution.[29] It had never sued OpenAI, whose chatbot leaned on Wikipedia more than any other source.[12] It had never demanded that Meta release the trained model itself under the same license that governed the text Meta trained on. The Foundation's own staff had admitted, in a March 2024 email, that many of these companies "may not be compliant with the letter of the Creative Commons rules or the spirit of the licenses."[18]

It kept selling to them anyway.

There was an uncomfortable irony buried in the outrage. In purely structural terms, Musk's copy was *more* compliant with Wikipedia's license than the companies the Foundation served. It credited Wikipedia and offered free public access.[25,31] OpenAI did neither, and paid nothing.[19] Grokipedia's neo-Nazi sourcing was a real and serious failure of content, but a failure of the visible kind that scrutiny can force out, and within months it mostly had: later versions scrubbed the most notorious citations.[26,34,35] The deeper harm was the one no one was protesting: the quiet absorption of Wikipedia into proprietary systems that could not be examined, corrected, or forked (freely copied and rebuilt) at all.

That harm had a history. In 1982, a company called Symbolics took code Richard Stallman had helped write, improved it, and refused to share the improvements back. He answered by inventing copyleft: a license

built to keep free work free.[2] Anyone who improved the work had to share their changes under the same terms.[5] Wikipedia launched in 2001 under a license of that kind.[36] The promise was simple: use this freely, but keep it free.

Twenty-four years later the promise had eroded, and not mainly at Musk's hands. OpenAI, Anthropic, Google, and Meta absorbed Wikipedia into proprietary models, gave no attribution to end users, charged for access, and licensed their outputs on their own terms.[7,8] Community work absorbed, proprietary value added, others locked out: Symbolics again, at industrial scale.

The institution built to prevent this collected more than $180 million a year.[23] It employed some 700 staff and contractors.[37] It ran a commercial arm, Wikimedia Enterprise, that sold structured data access to Google, Amazon, and Microsoft.[38]

And when the moment came to defend the commons its volunteers had built, it issued a press statement. The statement noted that "no single individual, company, or agenda can exert influence over the work."[27] It did not threaten legal action. It did not assert a licensing violation. After four years of accommodating the invisible extraction, it had forfeited the credibility to single out the visible fork.

Which brings us back to the man who named it.

On June 22, 2026, an administrator blocked Larry Sanger's Wikipedia account indefinitely.[39,30] Sanger had co-founded the encyclopedia in 2001 and given it its name.[36] No court ordered the block. No government, no corporation, no hostile takeover. Wikipedia's own community did it, through its own machinery: a noticeboard thread, a tallied "consensus," an administrator enacting it with a single click. "All my judges," Sanger wrote that morning, "were self-selected and hated me."[40]

The stated charge was off-wiki canvassing.[39] Sanger had used his following on X and other public platforms to pull outsiders toward an internal Wikipedia vote.[30] Wikipedia's rules forbid recruiting an outside audience into its consensus, and the community enforced them. The breach was real, and the rule is a reasonable one. But strip away the procedure and the substance is plain: Sanger was removed for *disagreement*, for attacking the project's direction for years and organizing, at last, to change it.[41]

He had copied nothing, enclosed nothing, broken no one's license. The companies the Foundation served had done all three. They breached the letter and the spirit of the copyleft license that was supposed to keep Wikipedia free, a rule the whole commons rested on, and they remained paying customers.[18,19] The dissenter lost his account. The license-breakers kept theirs.

Faced with a powerless critic who disagreed, the institution found the will to act, decisively, within days. Faced with the most powerful companies on earth breaking the license that defined it, it found only words. It had learned to discipline dissent inside its walls and to bill the forces dismantling it from without.

The consequences were already looping back. Grokipedia copied Wikipedia and injected its poisoned citations, and a newer AI model was soon caught citing that contaminated fork as a source on historical questions.[32,28] The human curation that made Wikipedia trustworthy, the careful sourcing and the blacklists and the community review, was being laundered out step by step, through the clean formatting of a chatbot's answer.

But this is not a story about Elon Musk, or even about artificial intelligence broadly. It is a story about one of our most important institutions. A volunteer movement to free the world's knowledge became a nine-figure foundation that sold access to that knowledge's extractors, then would

not defend the commons when it mattered.[23,38] The promise was broken, not by an outside force, but by the institution meant to uphold it. The antagonist is not a person. It is a failure: of nerve, of independence, of the will to enforce a promise.

And beneath the failure of will lay a failure of distance. Wikipedia was captured twice. The first capture was internal and nearly invisible: a political monoculture so complete that the Foundation kept no dissent, no balance, no distance from the powerful. It had grown alike in worldview, networks, and politics to the companies whose compliance it alone was positioned to demand. The homogeneity was real, and, as the public record shows, measurable to the dollar.[22]

The second capture followed from the first. Having lost the independence to see the AI industry as anything but partners, the Foundation could not resist it when it came for the commons. Stallman built copyleft to keep the enclosers out; he had not planned for a steward that would decline to enforce it.

The promise began in January 2001, with two founders, an argument about expertise, and a wiki that was never supposed to become the encyclopedia.[36] One of those founders would spend the next quarter century watching the project drift from everything he believed it should be. In the summer of 2026, it cast him out.[30]

Part I

Foundations and Fractures

CHAPTER ONE

The Expert Failure

> Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.[42]
>
> — Jimmy Wales, 2004

In September 2000, after months of drafting, review, and revision, an online encyclopedia called Nupedia put the finishing touches on an article about the musical concept of atonality. The author, Christoph Hust, had written something any academic journal would accept: 14 footnotes, 27 references, a complete discography of atonal compositions.[43] The prose was careful, the scholarship thorough. By every measure of traditional expertise, it was exactly what an encyclopedia should contain.

This was the dream of Nupedia, the free encyclopedia that Jimmy Wales and Larry Sanger had launched six months earlier.[36] Expert authors would produce expert content. A rigorous seven-step peer review process would ensure quality. The articles would be free to read, yes, but they would carry the weight of credentials. The internet would finally have an encyclopedia worthy of the name.

There was just one problem. After half a year of operation, Nupedia had produced exactly two finished articles.[3][6][1] The "Atonality" article was one of them.

Wales's company Bomis would eventually sink roughly $250,000 into Nupedia.[44] The economics were absurd. The pace, one article every three months, was worse. The internet was growing at a rate that demanded thousands of articles, not dozens. The experts weren't showing up. The process wasn't scaling. The vision was dying.

The bottleneck was not a temporary stumble. It exposed the flaw at the center of Nupedia's design: a system built to guarantee quality through expert review had guaranteed paralysis instead.

The Predecessors

The idea of a collaboratively written, freely available encyclopedia predated Nupedia by nearly a decade. It just never worked.[45]

In 1993, a computer scientist named Rick Gates proposed something called Interpedia. The concept was simple: volunteers would write encyclopedia articles and share them over the early internet.[45] Over 700 messages accumulated on the discussion mailing list. Fewer than 50 draft articles ever materialized. The project died before it could demonstrate whether collaboration at scale was even possible.[45] The tools didn't exist. The culture wasn't ready.

Six years later, Richard Stallman, the MIT programmer who had pioneered the free software movement, published a proposal of his own.[2] He called for a universal free encyclopedia built on the same principles as free software. His proposal supplied the copyleft premise Wikipedia would later inherit.[46]

[1] Article counts vary across sources. Secondary histories report 21 approved articles in Nupedia's first year;[4] Sanger's 2005 memoir recalls "only about 25" approved articles by early 2001.[36] The discrepancy may reflect different definitions of "complete" versus "in progress."

The project that grew from Stallman's essay, GNUpedia, launched in January 2001, the same month as Wikipedia.[45] The timing doomed it. Nupedia already existed, and Wikipedia, the fast-moving project that spun out of it, absorbed exactly the volunteer energy GNUpedia needed. Over the course of 2001, the GNU project stood down in Wikipedia's favor.[47]

The pattern was clear. Everyone agreed the internet needed a free encyclopedia. No one had figured out how to build one.

Bomis

Jimmy Wales came to encyclopedias through an unusual path. In 1996, he and two partners, Tim Shell and Michael Davis, founded a web company called Bomis. The site functioned as a web portal, helping users find content in the chaotic early internet. Some of that content was adult material, and "Bomis Babes" generated enough advertising revenue to keep the company afloat.[44]

Wales's intellectual interests, however, ran in different directions. As a finance undergraduate at Auburn University in Alabama, he studied with Mark Thornton, a scholar of the Austrian school of economics. Thornton, by Wales's own later account, got him to read Friedrich Hayek's work, particularly a 1945 essay titled "The Use of Knowledge in Society."[48,49]

Hayek's argument would shape everything Wales built. The essay contended that economic planning fails because the knowledge required for rational decisions is never concentrated in one place. It exists scattered across millions of individuals, each with their own fragment of understanding. No committee, however expert, can aggregate this dispersed knowledge efficiently. Only decentralized systems, markets in Hayek's view, can coordinate what individuals know into coherent outcomes.

Wales also absorbed Eric Raymond's 1997 essay "The Cathedral and the Bazaar," which applied similar reasoning to software development.[50] Raymond contrasted the "cathedral" model of software, carefully planned by a small team of experts, with the "bazaar" model of open source. In the

bazaar, anyone could contribute, and the best code emerged from many competing attempts. "Given enough eyeballs," Raymond famously wrote, "all bugs are shallow."

By October 1999, Wales had begun conceptualizing an online encyclopedia.[36] The question was how to build it.

Sanger Arrives

Larry Sanger had known Wales since the mid-1990s, when both argued philosophy on the era's mailing lists. By late 1999, Sanger was finishing his PhD in philosophy at Ohio State University.[36] His specialty was epistemology—the study of knowledge itself. How do we know what we know? What distinguishes justified belief from mere opinion? These questions had occupied philosophers for centuries. Sanger approached them with academic rigor.

In January 2000, Wales hired him to develop and run Nupedia. Sanger arrived in San Diego the following month to begin work.[36]

The philosophical tension between them was present from the start. Wales believed that good processes could substitute for credentials. Sanger believed that credentials signaled reliability precisely because knowledge requires justification, and justification requires expertise. Wales trusted emergence. Sanger trusted verification.[36]

For Nupedia, Sanger designed a process that reflected his training. If knowledge required justification, then articles would need to be justified—by experts, through review, at every stage. The result was a seven-step gauntlet that looked exactly like academic peer review because, in Sanger's mind, academic peer review was how knowledge got validated.[36]

Neither founder treated the difference as urgent. For the moment, they were building the same encyclopedia.

Seven Steps

Sanger's review system required seven sequential stages. First, a topic was assigned to a qualified author—ideally someone with a PhD in the relevant field. Second, an expert reviewer was recruited, also requiring a PhD "with few exceptions." Third, the lead reviewer checked accuracy and completeness, provided substantive feedback, and engaged in extended discussion with the author through a web forum. Fourth, other experts reviewed and commented, incorporating multiple perspectives in what Sanger compared to the question-and-answer portion of an academic conference. Fifth, a lead copyeditor improved grammar, punctuation, and layout. Sixth, additional copyeditors contributed refinements. Seventh and finally, the article received formal approval and was prepared for publication.[36]

The structure made sense if you assumed that encyclopedia articles should work like academic papers. Each stage guaranteed scrutiny. Each reviewer added legitimacy. By the time an article cleared all seven steps, it had earned the right to be trusted.

The problem was that each step blocked all subsequent steps. Nothing could move to stage three until stage two finished. Nothing could move to stage four until stage three finished. The system had no parallel processing. A single delay at any stage halted everything downstream.

And delays were constant.

The Absent Experts

The seven-step process assumed that qualified experts would volunteer their time. They didn't.[36]

The reasons were structural, not personal. Academic culture rewards publication in peer-reviewed journals, which counts toward tenure and promotion. Encyclopedia writing does not. A professor who spent hours reviewing a Nupedia article would have nothing to show for it profes-

sionally. The work wouldn't appear on a CV. Colleagues wouldn't cite it. Tenure committees wouldn't care.

Nupedia offered no compensation. The project depended entirely on intrinsic motivation—the satisfaction of contributing to something worthwhile. But intrinsic motivation works differently for tenured professors than for hobbyists. Professors already have outlets for intellectual contribution. Hobbyists often don't. The people most qualified to write expert encyclopedia articles were precisely the people with the least incentive to do so.

The committee structure made things worse. Academics are used to working alone or in small collaborations. They're not used to submitting their prose to seven sequential stages of review by strangers. The process felt burdensome rather than validating. It demanded more effort than academic publishing while offering fewer rewards.

By the summer of 2000, Sanger later wrote, it had become obvious to him and Wales that the seven-step process was moving far too slowly.[36]

The closure figures confirmed the verdict. By the time Nupedia shut down, only 24 articles had cleared all seven steps, with well over a hundred still stalled somewhere in the pipeline.[36,51] At roughly $10,000 per finished article, an encyclopedia of useful size would have cost billions. At that pace, roughly one article a quarter, it would have taken centuries. Wikipedia, once launched, would produce hundreds of articles in its first two weeks.[47]

The Deeper Problem

Nupedia's failure was not only a failure of speed. It was a failure of theory.

Nupedia assumed that knowledge production works the way academic publishing works. A small number of credentialed experts create content, and peer review validates it. This model functions in academia because journals carry prestige that attracts authors and reviewers despite the lack of payment. Encyclopedia articles had no comparable prestige. The incentive structure was broken from the start.

Wales had absorbed a different theory. Hayek's argument about dispersed knowledge applied to more than markets. Centralized expertise fails not because experts are foolish, but because knowledge is genuinely scattered across millions of people. No single authority can gather all of it and make it coherent. Only a system that lets many knowledge-holders contribute at once can create order without central planning. This was the dream of a commons: not the absence of rules, but rules created collectively by those with a stake in the resource.

Applied to encyclopedias, the idea was radical. Maybe the problem was not finding enough experts. Maybe the problem was relying on experts at all. Knowledge might emerge from the coordination of thousands of non-experts, each contributing a small piece, with the aggregate exceeding what any committee could produce.

This was heresy to Sanger's epistemology. Knowledge requires justification. Justification requires expertise. Without credentials, how could you trust what someone wrote?

Hayek's model suggested a different kind of trust: trust in process rather than credentials. If enough people contributed, and if good processes corrected errors, then the quality of any single contributor mattered less than the scale of participation.

The question was whether such a process could exist. In January 2001, Sanger stumbled across one.

The Dinner

On January 2, 2001, Sanger met his friend Ben Kovitz for dinner at a Mexican restaurant in San Diego. Kovitz was a computer programmer Sanger knew from online mailing lists. They ordered food. Sanger had enchiladas.[41]

Kovitz had been spending his spare time on something called WikiWikiWeb, a website created by programmer Ward Cunningham in 1995. The name came from the Hawaiian word for "quick"—Cunningham had

named it after the Wiki Wiki Shuttle bus at Honolulu International Airport. The concept was simple but strange: any visitor could edit any page, and changes appeared immediately.[52] There were no gatekeepers, no review stages, no credentials required.

Cunningham's site had grown into a thriving community of software developers. They wrote about programming patterns, debated design choices, and refined each other's contributions. When someone made an error, someone else fixed it. When vandals appeared, the community reverted their changes. Against all intuition, the system worked. The wiki self-corrected.[52]

Sanger listened with growing interest. "At some point I said, 'You know, I just wonder how this might solve the problem I've been having with Nupedia,'" he later recalled. "I was skeptical, obviously, because it doesn't sound plausible at first, but Ben was able to answer a lot of my objections. The best part of it was that the software was easy to install."[41]

That evening, Sanger and Kovitz went to Sanger's apartment. Sanger wrote a formal proposal and sent it to Wales. The proposal suggested using wiki software to accelerate content creation for Nupedia—a "feeder" system where rough drafts could be developed quickly before entering the seven-step review.[36]

Wales approved the experiment.

Kovitz never claimed credit for what followed. He had simply carried an idea from one community to another—a programmer who knew about wikis, having dinner with an editor who needed one.

The Rejection

On January 10, 2001, Clifford Adams's UseModWiki software was installed on the Nupedia servers. The wiki section launched as a test, positioned as a drafting area that would feed content into the main review process.[36]

The Nupedia Advisory Board was not pleased.

These were credentialed scholars—the very experts Sanger had recruited to give Nupedia legitimacy. They looked at the wiki and saw chaos. Anyone could edit? No review before publication? No verification of credentials? This wasn't an encyclopedia. This was a threat to everything Nupedia stood for.

"A clear majority of the Nupedia Advisory Board wanted to have nothing to do with a wiki," Sanger recalled.[36] Their commitment, one he shared, was to rigor and reliability.

Their objections were not irrational. Academic knowledge production had developed peer review for good reasons. Errors get caught before publication. Cranks get filtered out. Quality emerges from scrutiny.

A wiki inverted all of this. Errors would be published first and corrected later (maybe). Credentials wouldn't matter. Quality would depend on the crowd, and crowds are notoriously unreliable.

Sanger, though he had proposed the wiki, understood their position. "They were perfectly reasonable to doubt that it would turn into the fantastic source of content that it did."[36]

The Board's rejection forced a decision. The wiki couldn't be part of Nupedia. It would have to be something else.

Wikipedia Launches

On January 15, 2001, the wiki launched as a separate project under a new name: Wikipedia. Wales typed "Hello, World!" as the first edit.[2][53]

The name came from combining "wiki" with "encyclopedia." The domain was wikipedia.com (later changed to wikipedia.org to emphasize its nonprofit nature).[47] The software was UseModWiki, the same program installed for the Nupedia experiment five days earlier.[54] The licensing was the GNU Free Documentation License (GFDL), a copyleft license from Stallman's free-software world.[5]

[2]The original edit was wiped almost immediately when the site was reset; the earliest surviving edits come from later backups. Wales auctioned a re-creation of the moment as an NFT in 2021.

And the articles started appearing.

Within days, Wikipedia had more content than Nupedia had produced in a year. By the end of January, roughly 600 articles existed.[47] Contributors arrived from nowhere—hobbyists, enthusiasts, anyone curious enough to click "edit." They wrote about their interests, corrected each other's errors, and built something that no committee could have planned.

But this same openness, the feature that made Wikipedia powerful, would eventually invite extraction. An encyclopedia written by anyone could be read by anything. The copyleft license set conditions on reuse, but conditions protect nothing unless someone enforces them.

The self-correction Kovitz had described was real. When someone contributed something valuable, others refined it. When someone contributed something wrong, others fixed it. When vandals appeared, they were reverted within minutes. The process wasn't perfect, but it was fast. And speed, it turned out, mattered more than perfection.

By January 2003, the English edition alone had passed 100,000 articles.[47] The experiment had become the project.

The Nupedia Advisory Board had grasped the stakes correctly. The wiki was not just a technical change but a philosophical break. Nupedia treated knowledge as something experts produce and gatekeepers validate. Wikipedia treated it as something that emerges from process, from the contributions of thousands of people, most anonymous, none credentialed. The Board saw that these were incompatible visions, and that a wiki would undermine Nupedia's whole premise. They misjudged only which model would win.

What the Board did not anticipate was that the internet had changed the economics of knowledge production. Nupedia needed experts who would work for free despite no professional incentive to do so. Wikipedia tapped a different pool: people who wanted to build something larger than themselves, who had passion for their subjects, who found satisfaction in a shared resource. These people existed in the millions. Academics with

PhDs existed in the thousands. Scale beat credentials, and almost no one saw it coming, Sanger included.

The Aftermath

Nupedia lingered for two more years. After 2001, only two more articles completed its seven-step review.[51] The community's energy had shifted to Wikipedia, where a single contributor could write and publish in an afternoon what Nupedia took months to produce. The expert model couldn't compete.

On September 26, 2003, Nupedia's servers were shut down for good.[51] By then, Wikipedia had more than 150,000 articles in English alone, created by contributors whose only credential was the willingness to edit.[47]

Sanger had left the projects 18 months earlier, in March 2002.[55] But the philosophical battle was already lost by the time he left. Wikipedia had chosen Wales's model over Sanger's. The community had voted with their keystrokes, and they voted for velocity over verification.

The original idea had been an encyclopedia written by experts, for everyone, with the wiki serving as its feeder system. Instead, the feeder ate the host.

The Founding Choice

Christoph Hust's "Atonality" article, with its 14 footnotes and 27 references, was the last gasp of a model that valued credentials over contribution.[43] By March 2002, anonymous contributors had written their own article on atonality.[56] Less scholarly, perhaps, but read by millions.

The choice Wikipedia made in January 2001 had consequences that reach the present. They followed from a single fact: the content carried no credentialed author.

Because no expert stood behind any sentence, Wikipedia was hard to defend as proprietary. When AI companies later scraped it for training data, they took content whose creation had privileged no single source of

authority.[18] The openness that defeated Nupedia's gatekeeping became, two decades on, the opening through which anyone could extract. And because the work was done by volunteers, the project had no employment contracts, no union, no collective bargaining power to set against commercial taking. Mass participation made Wikipedia resilient to any one person leaving. It also left the commons exposed to systematic extraction, and made any particular harm hard to prove. The project had chosen accessibility over defensibility.

One consequence emerged more slowly. Wikipedia's anti-credentialist philosophy said nothing about who would staff the Foundation built to steward it. That would turn on hiring networks, geographic sorting, and cultural fit, and those forces would produce a homogeneity of their own—one that would, decades later, become measurable in the public record. An institution composed of like-minded people can move quickly and agree easily. It can also lose the capacity to recognize, when extraction finally arrives, that the people arriving to extract are not partners.

The founding tension, between openness and control, between process and credentials, was never resolved. The Spanish-language editors who broke away in 2002, fearing commercial control, fractured over these questions first.[57] Sanger's departure made the philosophical split permanent. And when AI companies began taking Wikipedia wholesale, the openness its founders celebrated offered no resistance. The license did, but only if someone chose to enforce it.

The question was no longer whether the open model could work. It had worked spectacularly. The question was what it would cost, and who would pay.

CHAPTER TWO

The Spanish Fork

> The international Wikipedia that you all know and have come to take for granted might have been impossible without the Spanish fork.
>
> — Edgar Enyedy, 2011

In mid-February 2002, about two weeks after Jimmy Wales told Wikipedia's mailing list that the site would soon carry advertising, Edgar Enyedy sat down to write a farewell message. For six months he had devoted himself to building the Spanish Wikipedia. Eight to twelve hours a day, every day. He created articles, developed a style guide, and designed an index system based on the Universal Decimal Classification, a library cataloguing scheme. He had a master's degree in communications systems and networking from the Polytechnic University of Madrid, training in philology and computer science, a background in journalism. He had poured all of it into Wikipedia.[57]

Now he was leaving.

The message itself has not survived. But its parting line, as Enyedy later recalled it, became legendary in Wikipedia's early history: "Good luck with your wikiPAIDia."[57]

The neologism captured everything wrong in a single word. Wiki plus PAID. Free knowledge corrupted by money.

But the anger went deeper than advertising. Enyedy had volunteered thousands of hours to build Spanish Wikipedia, believing he was contributing to a commons, a shared resource governed by and for its contributors. In fact Wikipedia ran on the servers of Bomis, Jimmy Wales's for-profit company, and under its control.

Advertising was how a site like that turned readers into revenue, and Wikipedia had readers only because volunteers like Enyedy had written the articles that drew them. The content, and the audience it pulled in, were worth money, and that value would flow to Bomis, not to the people who built it. His free labor had been building Bomis's asset. In the language of commons governance, this was enclosure, the appropriation of shared resources by a private owner.

Enyedy wasn't alone. When he announced the fork, the launch of a separate, independent copy of the project, questions flooded in: "And now? What's next?" The answer was Enciclopedia Libre Universal en Español, the Free Universal Encyclopedia in Spanish. A new project, hosted at the University of Seville, free from commercial control forever.[57,58]

Most of the Spanish Wikipedia's active contributors followed him out the door.[4]

First Success

The Spanish Wikipedia launched in May 2001, just four months after the English version.[57][1] By early 2002 it was one of the largest non-English Wikipedias.[4] It was proof that the wiki model could work across languages and cultures.

Enyedy and Javier de la Cueva, a lawyer specializing in technology and internet issues, emerged as its unofficial leaders. They didn't just write articles. They built infrastructure: policies, style guides, organizational systems. They communicated with Wales and Sanger in San Diego about

[1] Sources disagree on the exact date: the Spanish Wikipedia's own records point to May 11, 2001, while the Wikimedia Foundation's anniversary materials use May 20.[59]

the challenges facing international Wikipedias. They tried to make the project work.[57]

But the challenges were substantial, and many had nothing to do with content.

The Grievances

Wikipedia was hosted by Bomis. All the domains, wikipedia.com, wikipedia.org, and wikipedia.net, belonged to Wales personally. The servers sat in San Diego. Bomis staff controlled the software updates.[57,36] No formal governance structure separated editorial decisions from commercial interests.[45]

For the English Wikipedia, this ambiguity was tolerable. The English editors worked closely with Wales and could negotiate directly. For international Wikipedias, the arrangement felt colonial. They built content for an American company, subject to decisions made in a language many of them didn't speak, with no formal say in the project's direction.[57]

Enyedy later enumerated his grievances. They went far beyond advertising.

"I asked myself 'why are we working for a dot-com?'" he recalled in a 2011 interview. "I asked for Wikipedia to be changed to a dot-org."[57]

He objected to Larry Sanger's role as editor-in-chief: "I wanted the Big Brother out. Larry Sanger was against the nature of the project itself. None of us felt comfortable with such a figure."[57]

He pushed for autonomy: "We did not want to be seen as mere translations of the American version... I was told so many times to translate from the main wiki, and my response was always the same: We are not a translation of the American Wikipedia!"[57]

The governance complaints had concrete proof. Software updates went to the English Wikipedia first. International versions ran outdated code for months. When Enyedy requested access to the Spanish server to help maintain it, his request was denied. "They said it was for security

reasons because Bomis Inc. was hosting files from its clients on the same server."[57]

He asked for mirror servers, backup copies that would ensure the Spanish content survived independently. The answer was always the same: the project needed to stay together. Database exports, which would have let contributors copy their work elsewhere, weren't being updated. The license that governed Wikipedia's content, the GNU Free Documentation License, promised that anyone could fork the project and host their own version.[5] In practice, the technical barriers made forking nearly impossible.[57]

"The Wikipedia page on Sourceforge had instructions that read like hieroglyphics," Enyedy said. "And once again due to 'technical' reasons that none of us believed, the downloadable database was never updated... These conditions did not resemble what the GNU/FDL was supposed to ensure."[57]

Here was the betrayal beneath the surface. Copyleft was the shared resource's insurance policy. If Wikipedia ever became proprietary, anyone could fork it. But Bomis controlled the servers, and server control determined whether the legal right to fork meant anything in practice.

The Fork

In late 2001, the internet bubble collapsed. Bomis lost major contracts. Revenue dried up.[60] By December, Tim Shell, a Bomis co-founder, had told Larry Sanger he would be laid off. The layoff took effect in February 2002. Wikipedia was costing money Bomis no longer had.[36,60]

On February 2, 2002, Jimmy Wales posted to the Wikipedia-L mailing list: "However, with the ongoing hard times in the Internet economy, we do anticipate adding some forms of advertising to the site in the near future."[61]

Sanger, now working unpaid, mentioned that advertising revenue "might make it possible for me to come back to my old job."[4]

For Enyedy, this was confirmation. Everything he feared about Bomis's commercial control was coming true. The announcement was not a surprise. It was the last straw on a pile of grievances months in the making.

"The possibility of advertising was out of the question," he later said. "I asked Wales for a public commitment that there would be no advertising. This only came after we left."[57]

On February 26, 2002, Enciclopedia Libre went live.[62] Juan Antonio Ruiz Rivas had arranged hosting at the University of Seville: academic, non-commercial, permanently outside Bomis's control.[57]

The GFDL made forking legally possible.[5] The practical barriers remained. The database exports weren't current. There was no automated way to migrate content. The Spanish contributors had to copy their articles one by one.[57]

"When the server was up and running, and as the GNU/FDL permitted, we began copying our articles from Wikipedia," Enyedy recalled. "It wasn't an automated process, no bots or anything, just us bringing the articles across one by one from Wikipedia's server to ours."[57]

The exodus was swift. Enyedy "persuaded most of the volunteers to go with him." The Spanish Wikipedia, one of the most active non-English versions, was gutted. It stayed nearly dormant into 2003. "It took more than a year for the Spanish Wikipedia to get back on its feet again," Andrew Lih wrote in his history of Wikipedia.[4]

Wales Responds

On March 1, 2002, Larry Sanger resigned as Wikipedia's chief organizer. The same day, Wales sent an email to Wikipedia-L with a significant reversal:

"With the resignation of Larry, there is a much less pressing need for funds. Therefore, all plans to put advertising of any kind on the wikipedia is called off for now."[63]

The qualifier, "for now," was notable. Wales continued: "Just as the National Geographic Society is supported in large part by advertisements in the National Geographic Magazine, I expect this to be a potentially necessary thing at some point in the future, if we wish to have an impact beyond our own little corner of the Internet."

Even in 2006, Wales told the marketing outlet ClickZ: "We have never said there would absolutely never be ads on Wikipedia."[64]

For the immediate crisis, the concession was enough. Advertising was off the table. The Spanish forkers had won their first demand.

In August 2002, Wikipedia changed its domain from .com to .org, signaling non-commercial intent.[47] In October, a Wikipedian named Daniel Mayer, known as Maveric149, led a reunification effort, hoping to bring the Spanish contributors back. The Enciclopedia Libre community held a "Retorno a Wikipedia" vote. They rejected reunification.[65]

On June 20, 2003, the Wikimedia Foundation was incorporated as a nonprofit corporation in Florida.[66] The IRS granted 501(c)(3) tax-exempt status in April 2005.[67] Wikipedia was formally separated from Bomis. A foundation would govern the project, not a company.[66]

The fork came first. The concessions followed. Within sixteen months, Wikipedia had a nonprofit foundation between the volunteers and Bomis.[66] Enyedy was right to claim credit. Wales feared other language communities would follow the Spanish example. He learned, as Enyedy put it, "what to do and what not to do."[57]

Founder Dispute

Years later, the advertising controversy generated its own controversy. In January 2011, when Wired UK published a story about the Spanish fork, Wales wrote in the comments section:

"Sanger was absolutely adamant that Wikipedia must have ads, and it was my refusal to do so that led to Wikipedia being as it is today. The

Spanish fork did not provoke any changes of any kind. We stayed the course."[60]

Sanger, who had long since broken with Wales,[41] responded on his blog:

"The suggestion that I demanded ads and that Jimmy Wales was opposed to them is, I am afraid, yet another self-serving lie from Wales... From the beginning, Wales let me know in no uncertain terms that, once it garnered enough traffic, Nupedia would become ad-supported."[60]

Sanger cited Wales's February 2, 2002 email as evidence. He noted that Wales had deleted a Wikipedia discussion where editors pressed him on the discrepancy between his 2011 claim and the historical record.[60] That deletion was possible only because no rule yet stood in its way.

The mailing list archives contradicted Wales's later account.[61] He had announced advertising plans. The community had revolted. He had reversed course. These were documented facts.

But the dispute revealed something deeper. The founders themselves couldn't agree on what they had intended.[60] No governance structure existed to settle such questions. Everything depended on the memory and goodwill of the people who controlled the servers.

The Spanish contributors fought about advertising because advertising was visible and concrete. The underlying problem remained even after the Foundation was created: no governance mechanism constrained central authority.

To Enyedy and the Spanish forkers, the revolt was about principle, not convenience. Volunteers should not build value for a for-profit company without governance guarantees. A license that promised freedom meant nothing if technical barriers prevented forking.

"We were basically working for Bomis Inc., and asked in a gentle way to translate from the main Wikipedia," Enyedy summarized. "Finally, came

the possibility of incorporating advertising, so we left. It couldn't be any other way."[57]

From Wales's perspective, and the official Wikimedia Foundation history, the fork was a misunderstanding resolved. Wikipedia was always meant to be ad-free. The February 2002 announcement was an aberration caused by financial pressure, not a statement of intent. The Foundation's creation reflected natural evolution, not forced concession. The project "stayed the course."[60]

Both accounts hold a piece of the truth, and both miss what Enyedy had actually seen. The fork was not only about advertising. It exposed a pattern that would define Wikipedia's next two decades. A shared resource stays stable only while its contributors believe they control it. The moment a central authority takes it for profit, the social contract collapses.

The forkers were right about the threat. They were wrong only about its form. They feared advertising. They could not have imagined the Foundation one day selling bulk access to its content for AI training.[29] Nor could they anticipate a subtler capture: the institution created to protect against commercialization would come to be led by people who saw tech companies as natural allies, not adversaries. They won concessions against banner ads, the visible symbol of commercialization.[63] They established no protection against extraction through training data.

Enyedy's 2011 warning proved prescient: "This is what I didn't want to happen: a large, money-centred organisation made possible by the free work of the community."[57] When he said this, the Foundation's annual revenue was about $25 million. By 2024, it exceeded $180 million.[23]

In 2011, the scholar Nathaniel Tkacz interviewed Enyedy for an academic volume on Wikipedia. The forker offered a clear-eyed verdict on what he had accomplished:

"The fork had its time and place, its goal and its consequences. Nowadays, the romantic point of view is that EL survived and is still going strong. It is a nice view, but wrong. EL has failed as a long-term project

for one reason: The project itself was not intended to last. It was merely a form of pressure. Some of the goals were achieved, not all of them, but it was worth the cost."[57]

The goals achieved were real. Wikipedia became ad-free, moved to .org, established the Foundation, gave international editions more autonomy, improved its database exports, and put its software on SourceForge.[57] The Spanish Wikipedia recovered and grew to millions of articles.[68] In late 2024, with no announcement anyone noticed, Enciclopedia Libre ceased operations.[62] Twenty-two years after its founding, the fork that almost was ended with a whisper.

By then the fears that motivated it were no longer hypothetical. Wikipedia never ran advertising.[47] It had become instead the training data for a new generation of commercial products,[69] and the Foundation created to protect it was taking the money.[20]

The volunteers won on the threat they could see. The threats they could not see arrived later. That was the gap the fork exposed and the Foundation never closed: no mechanism to constrain whoever controlled the servers. It would outlive the people who first noticed it.

CHAPTER THREE

The Departure

> As a community, Wikipedia lacks the habit or tradition of respect for expertise.
>
> — Larry Sanger, 2004

Larry Sanger explained his departure twice, and the two explanations did not match.

His formal resignation, posted on March 1, 2002, said all the right things: no loss of faith in the project, a relationship with Jimmy Wales and Bomis that "remains friendly," a departure that was purely financial. The dot-com bubble had burst, and Bomis could no longer afford his salary.[70,55]

This was true as far as it went. But the fuller confession, written months later on Meta-Wiki, the community's side site for discussing the project itself, told a different story:

"I must now confess, when I had to quit as 'organizer' of Wikipedia at the end of February 2002, that it was a relief."[55]

A relief. The only paid Wikipedia employee. The man who named the project, designed its policies, served as its chief organizer for its entire first year.[36] And leaving felt like escape.

"Very much of my time was spent in defending myself from such attacks," he continued. The attacks came from within, from other Wikipedians who opposed his vision of how the project should work. Certain

critics, he wrote, had mounted "a concerted effort to criticize and undermine virtually every 'management'" action he took.[55]

The wiki structure worked against him. "Critics in a wikiwiki environment cannot be silenced... no criticism, no matter how false or misleading, can be erased without an accusation of censorship." Every dispute escalated. Every editorial decision became a referendum on his authority. By the time Bomis informed him the funding would end, the poisonous atmosphere had already driven him to the brink of quitting.[55]

What happened in those final months? What tension could make the creator of Wikipedia's policies desperate to leave his own creation?

The Partnership

In January 2001, when Wikipedia launched, Larry Sanger was its only paid employee. Jimmy Wales provided funding through Bomis and made high-level decisions, but Sanger was the one building the project day by day.[4] He wrote or commissioned the first articles. He formulated the Neutral Point of View policy that would become Wikipedia's philosophical foundation. He moderated disputes, recruited contributors, and shaped the community's norms.[36]

The press releases from those early months credited both men as founders.[71] In September 2001, the New York Times reported that Sanger had "founded Wikipedia with Mr. Wales," a characterization Wales disputed for years afterward.[72,73] The relationship appeared collaborative. Two men with different skills, building something together.

But the philosophical tension between the two men had never resolved. Wales believed in distributed knowledge, emergent order, process over credentials.[45] Sanger believed expertise should have standing, that credentials signaled reliability, that quality required some form of gatekeeping.[74] Wikipedia had chosen to be open. Sanger kept trying to impose structure on that openness.

For a while, both visions coexisted. Then they didn't.

Atmosphere Curdles

By fall 2001, something had changed. Wikipedia was growing explosively, with 20,000 articles by year's end, but Sanger's authority was shrinking.[47,55]

The attacks weren't coordinated, but they shared a common theme: Sanger was too controlling. His attempts to maintain standards were authoritarianism. His editorial judgments were elitism. Every time he tried to enforce quality, someone accused him of violating Wikipedia's open culture.[55]

The structural problem was that Wikipedia had no mechanism to resolve such disputes. In a traditional organization, Sanger's title as editor-in-chief would have settled things. In Wikipedia's radical flatness, titles meant nothing. Anyone could criticize anything. And in a wiki environment, criticism couldn't be removed without triggering accusations of censorship.[45,36]

"It's now generally the case that if I attempt to do or say anything that is very controversial—certainly anything regarding Wikipedia policy and politics—I am either ignored or, more often, vociferously denounced," Sanger wrote.[55] The community he had helped build was rejecting him.

Troll Problem

Wikipedia's openness attracted both contributors and disruptors. Sanger described a "terrible dilemma."[55] The same radical accessibility that powered Wikipedia's growth also empowered those who wanted to damage it.

"Trolls" in early internet parlance were users who disrupted communities for entertainment. They started fights, inserted false information, and wasted contributors' time with bad-faith arguments.[75] A traditional encyclopedia could simply ban them. Wikipedia's commitment to openness made that much harder.

Sanger tried to address the problem. He proposed giving experts more authority. He suggested mechanisms to identify and remove troublemakers. Each proposal met the same objection: you're betraying Wikipedia's principles. The project's most active contributors were "decidedly anti-elitist," he observed.[74] They saw his attempts to impose structure as attacks on the open model that was working.

The irony was bitter. Sanger had designed many of Wikipedia's policies. Now those policies were being used against him.

Financial Trigger

In late 2001, the internet economy collapsed. Bomis, which had been funding Wikipedia through advertising revenue, lost major contracts.[55] In December, Bomis told Sanger it could no longer afford his salary. He was one of several employees the crash swept out.[36]

Wales promised another push to sell advertising on Wikipedia to keep Sanger on staff.[60] That attempt led to the February 2002 advertising fight that drove Spanish-language editors to break away and start their own encyclopedia.[57] But for Sanger personally, the financial news was secondary. The community had already made his position untenable.

When the resignation came on March 1, 2002, the note tried to hold the project together. "My departure should not be taken as a reflection on Wikipedia, or you," Sanger wrote. "It still might succeed brilliantly."[70] That same day, Wales emailed Wikipedia-L and called off the advertising plans that had provoked the breakaway.[63] The forkers had won that fight. Sanger did not return.

Who Decides

The deeper problem outlasted the resignation. Even as a volunteer, Sanger found he could not contribute "either with regard to policy and (collective) management or to the improvement of articles, without myself feeling...

that at any thing I say, I will be forced into a completely ridiculous, unseemly, undignified flamefest."[55]

He had tried to institute "a policy of respecting and deferring politely to experts."[55] It failed. The community would not accept expertise as a basis for authority. The only authority was consensus, and consensus had turned against him.

His departure exposed the deepest question facing Wikipedia as a commons. Who has the authority to make decisions? In a traditional encyclopedia, the editors decide. In a market, the buyer decides. In a commons, everyone affected should in theory have a voice. But "everyone" was millions of anonymous people who contributed once and vanished.

Sanger had tried to make expertise the deciding principle. The community rejected it.[55] The rejection was not anti-intellectual. It was fiercely democratic, and democracy in a commons is messy.

The departure was not just from a job. It was from a community that had decided it did not want what he offered.

Community Response

How did the community react to losing its founding organizer? With muted relief.

Erik Moeller, later the Wikimedia Foundation's Deputy Director, wrote on Kuro5hin: "Larry Sanger played a crucial role in the first year of Wikipedia and has helped establish a shared philosophy for its users... This may well be the right time to abandon the concept of 'editorship' altogether."[76] The transition from editorial leadership to self-regulation was framed as progress, not loss.

Others on Meatball Wiki noted: "The project now no longer has a leader (or, put another way, everyone is a leader now)."[77]

In these public responses, at least, relief outweighed grief. The anti-elitist philosophy that Sanger had struggled against for months had won.

Wikipedia would be governed by process, not authority. By consensus, not expertise. By crowds, not editors.

The Divide

Sanger made the underlying conflict explicit in December 2004, in an essay on Kuro5hin. Titled "Why Wikipedia Must Jettison Its Anti-Elitism," it articulated the critique he had been developing since his departure.[74]

Experts would avoid Wikipedia, he argued, because their credentials carried no weight. "Nearly everyone with much expertise but little patience will avoid editing Wikipedia, because they will—at least if they are editing articles on articles that are subject to any sort of controversy—be forced to defend their edits on article discussion pages against attacks by nonexperts."[74]

He predicted the consequences. Trolls would flourish because the community wouldn't enforce standards against them. Experts would stay away. Wikipedia would become a source that knowledgeable people couldn't trust.

He named his expectation for the future: "as Wikipedia increases in popularity and strength, I do not see how there can not be a more academic fork of the project in the future."[74]

He acknowledged that none of this would change. "I certainly do not expect Jimmy Wales to change his mind. I have known him since 1994 and he is a smart and thoughtful guy; I am sure he has thought through his support of radical openness and his (what I call) anti-elitism."[74]

The founding schism wasn't a misunderstanding to be resolved. It was a fundamental disagreement about how knowledge should be organized. Wikipedia had chosen one model. Sanger represented the road not taken.

His critique aimed at quality. But the same openness he fought had a second edge: it made Wikipedia easy to copy, and it left no one charged with defending the terms of the license. The flat, anti-hierarchical model gave no one clear standing to speak for the project's value. Sanger had

wanted enforceable authority. The community refused it, and the question of who would defend its openness passed to an institution that did not yet exist.

Two Readings

How the departure was remembered split along the same line that caused it.

In the mainstream history, and in Wales's telling, Sanger's exit was a sad but ordinary casualty of the dot-com crash. Bomis lost money. Salaries were cut. Sanger left. Wikipedia continued and succeeded spectacularly.[4]

The evidence for that reading is substantial. Wikipedia grew from some 20,000 articles in early 2002 to over 350,000 by late 2004.[47] The project stayed the course on its founding principles and became the world's largest encyclopedia. Sanger's later alternative, Citizendium, peaked briefly in 2008 and then faded, seeming to prove that an expert model could not work at scale.[78]

Wales treated the co-founder dispute as settled and tiresome. In 2009 he called it "silly."[79] In 2023 he said Sanger "doesn't get enough credit for his early work" while holding that "co-founder's not the right title."[80] In November 2025, asked about it on a German podcast, Wales walked out after 48 seconds and called it "the dumbest question in the world."[81] The tension stayed raw, but the trajectory looked clear. Wikipedia won. Sanger lost.

Sanger told it differently. The financial trigger obscured the real cause, which was philosophical rejection. Wikipedia's egalitarian culture had made his role impossible, and the community would not accept his vision of structured expertise. Even as a volunteer, he found continued participation untenable.[55]

In his reading, Wikipedia's success came at the cost he had predicted. Citizendium failed, he argued, not because his model was wrong but because Wikipedia already held insurmountable network effects.

The name, the policies, a year as chief editor, the first articles written or commissioned: if that record did not make him a co-founder, he asked, what would the word even mean? He had also drafted the Neutral Point of View policy that anchors Wikipedia's editorial system to this day.[36] Yet the institution built to serve the community had no seat for him. The person who designed the governance was excluded from it.

The disagreement did not resolve. It went dormant. The question underneath it returned in every later conflict between the community and its institutions: who governs open knowledge, and should credentials matter?

What Followed

After March 1, 2002, Wikipedia continued without formal editorial leadership. Wales became the sole public face of the project.[4] The anti-credentialist philosophy was locked in. The Neutral Point of View policy Sanger had drafted remained, and the community built Verifiability and No Original Research on its foundation the following year.[82] His approach to enforcement did not survive.

The Spanish forkers fared no better than Sanger. They won concessions on advertising but not on structural change.[57]

The co-founder dispute festered for decades, escalating with each round. In 2005, Wales edited his own Wikipedia biography 18 times, seven of them altering information about Sanger's co-founder status. The next year he told the Boston Globe that Sanger's claims were "preposterous."[73] By 2009, in a public quarrel stoked by entrepreneur Jason Calacanis, Sanger was calling Wales a "fraud" and a "liar."[83]

Long Echo

In September 2025, Larry Sanger published a 37,000-word document, "Nine Theses on Wikipedia," an echo of Martin Luther nailing his complaints to the church door. He proposed competing articles on controversial topics,

an end to source blacklists, and mandatory editor identification.[84] The document was the culmination of two decades of critique, refined through failed alternatives and evolving grievances.

The very next day, Elon Musk shared the theses, "Some good suggestions from the co-founder of Wikipedia," and announced Grokipedia.[85] The Washington Post profiled Sanger as "inspiring Elon Musk to build a rival."[86]

Twenty-three years after his departure, Sanger's critique had found its most powerful amplifier, and a complicated validation. Grokipedia was an AI-built fork of Wikipedia, openly licensed and free to read[31] but salted with fringe sources egregious enough to make it notorious.[26]

But Sanger evaluated Grokipedia as he would any encyclopedia, by content quality, room for improvement, and structural openness.[87] He graded his own entry a "C" and called its confident AI fabrications "bullshittery."[88] He was not blind to the problems. He simply held that a flawed public encyclopedia beat no alternative at all, and that the freedom to fork was fundamental to the open project he had helped create.

The irony lay elsewhere. Grokipedia's content failures were visible and correctable, the kind a public fork invites and its critics can catch. The deeper extraction, Wikipedia absorbed into proprietary AI systems, went unseen and unchallenged.

The schism had not healed. It had moved into a new arena.

Cast Out

In June 2026, the community answered Sanger.

For twenty-four years Sanger had used voice, the dissenter's option of arguing from within. He critiqued, he built alternatives, he published the Nine Theses, and in 2026 he organized a WikiProject, an editor-run working group, devoted to intellectual diversity inside Wikipedia itself. He urged his followers toward the internal debate.[30]

The answer was exit, and not the voluntary kind. An administrator blocked his account indefinitely, carrying out a decision reached on the community's public administrators' noticeboard.[30] Editors charged him with off-wiki canvassing, meaning he had rallied outside supporters to sway an internal debate.[30] They also accused him of treating the project as a battleground. The procedural breach was real.

The verdict was neither smooth nor unanimous. A first block landed before the discussion's seventy-two-hour window had run. It was lifted, then reimposed once the window closed and the consensus was tallied.[30] The loudest objection came from the other founder. "I find the idea of any of this deserving an indef ban for Larry is ludicrous," Jimmy Wales wrote on his own talk page the morning of the final block, adding that editors needed to "sit back and have a hard look at what they are saying."[89] In 2002, Wales's judgment had settled Wikipedia's disputes. In 2026 it counted for one comment in the thread, and the thread went the other way.

But the timing closed an arc that had opened a quarter-century earlier. The co-founder who walked away in 2002, because he could not abide the project's openness, was in the end cast out by the project itself. He had once left on his own terms. This time the community decided for him.

CHAPTER FOUR

The Policies

> The threshold for inclusion in Wikipedia is verifiability, not truth.
>
> — Wikipedia policy, 2005–2012

In Wikipedia's first years, editors with a background in physics kept running into the same kind of problem.[90]

Picture a contributor who inserts a fringe energy theory into the article on cold fusion. An editor reverts it. The contributor restores it. Back and forth, week after week. The theory claims to explain a new form of energy generation, and the only source for it is the contributor's personal website.

An editor would object that a personal website was no source at all. The contributor would insist the theory was true and fully worked out, and that the absence of a journal article proved nothing.

Here was the bind. The physics editor knew the theory had no place in an encyclopedia, but Wikipedia had no formal policy to cite.[91] The editor could not invoke credentials, because Wikipedia had rejected credentials as a basis for authority.[74] The editor could not simply ban the contributor, because Wikipedia's openness let anyone edit.[47] The only argument available was that this was obviously fringe nonsense, and the contributor did not find it obvious at all.

The physics editors needed something beyond common sense. A standard. An external reference point that could settle disputes without requiring agreement on what was true.

This was the physics crank problem. Its solution became Wikipedia's working rule for knowledge.

Sanger Void

When Larry Sanger departed in March 2002, Wikipedia lost more than an employee. It lost the only person who had claimed authority to make editorial decisions.[55]

In February 2002, just weeks before his resignation, Jimmy Wales had written: "Larry is the final arbiter of what the consensus is."[92] After Sanger left, Wales inherited this role by default rather than design. He made decisions when asked, intervened in disputes when necessary, and served as the project's public face. But his authority was never formalized. Wikipedia had no constitution, no bylaws, no governance structure at all.[45]

This worked, for a while. Wikipedia was still small enough that disputes could be resolved through conversation. Wales was present enough to adjudicate when conversations failed. The community shared enough values that most questions had obvious answers.[45]

But the project was growing. The English edition passed twenty thousand articles in January 2002 and fifty thousand by early October.[93] More contributors meant more disputes. More disputes meant more need for clear rules. And some of those disputes, like the physics crank problem, could not be resolved by appeal to founder authority. They needed principles.

Scale Problem

The challenge of growth wasn't just quantity. It was the kind of disputes that scale produced.

When Wikipedia had a few hundred active editors, they knew each other. Disagreements could be resolved through reputation and relationship. A respected editor's opinion carried weight because other editors knew their track record. This was informal expertise: not credentials, but earned trust.[45]

As the community grew, this social knowledge became impossible to maintain. New editors arrived who knew nothing about existing contributors. They inserted content, disputed deletions, and demanded explanations. "Who are you to say this is wrong?" was a question that small communities could answer through familiarity. Large communities needed something else.

Disputes began to settle into a familiar pattern. Editors would argue: "I know this is true." Other editors would respond: "No, it isn't." Without a shared standard for evidence, such disputes could continue indefinitely. The physics cranks were the most dramatic example. People genuinely believed in perpetual motion or cold fusion and would fight endlessly for inclusion.[90] But the pattern appeared across domains. Political disputes, historical interpretations, medical claims. Everywhere, the same dynamic: my truth against yours.

Wikipedia needed a way to settle such disputes without resolving the underlying disagreement about truth.

Cultural Context

The solution to the truth deadlock reflected Wales's intellectual commitments: distributed knowledge, emergent order, process over credentials. He had built Wikipedia on the conviction that experts could not be trusted to adjudicate truth.[45] So perhaps the question itself could be reframed, shifting the ground from truth to something a divided community could share.

This was more than a procedural trick. It changed the question from what was true to what had been published. The project would become a curator of existing knowledge rather than a producer of new knowledge.

This reframing produced the policies that still govern how Wikipedia decides what it knows.

NPOV

The Neutral Point of View policy, universally abbreviated NPOV, was the oldest and most fundamental of Wikipedia's content policies.[45] Larry Sanger had drafted an early version for Nupedia in 2000, calling it the "Non-bias policy." He adapted it for Wikipedia by February 2001.[36] Wales elaborated the concept with what he later called his "original formulation" in April 2001.[94] Sanger wrote an expanded version at Meta-Wiki in December 2001.[36]

In November 2003, Wales declared NPOV "non-negotiable," a statement that simultaneously codified its importance and constrained future debate. "A few things are absolute and non-negotiable, though," he wrote. "NPOV for example."[95]

What did NPOV do? It was the first true commons governance policy. Instead of asking "What is true?" (which divides any community), NPOV asked "What do reliable sources say?" The reframe worked because it let people with incompatible worldviews work on the same articles. An atheist and a believer could both edit an article on God by agreeing to describe what philosophers and theologians have written. NPOV was not neutral about everything. It privileged published sources. But this particular partiality allowed a commons to function.

At its core, it required that articles represent all significant viewpoints fairly, without asserting that any particular viewpoint was correct.[96] Wikipedia would describe controversies rather than resolve them. If reliable sources disagreed, the article would present the disagreement. If

one view was dominant in published sources, the article would reflect that dominance while still noting alternatives.

The policy worked by deferring the hardest question. NPOV didn't require editors to agree on truth. It only required them to agree on what sources said. Two editors who disagreed about evolution could collaborate on an article if they could agree that published biology textbooks overwhelmingly supported evolutionary theory while creationist sources existed. The article would describe both, with appropriate weight, and neither editor had to abandon their personal beliefs.

Joseph Reagle, in his study of Wikipedia's culture, called this "productive ambiguity":

"While the perception is that NPOV is the source of much debate, it may act rather as a heat shield: reducing conflict and otherwise channeling outstanding arguments in the productive context of the primary goal of developing an encyclopedia that is representative of many viewpoints."[45]

But NPOV couldn't solve every problem. What made a viewpoint "significant"? How should conflicting sources be weighed? And critically: NPOV addressed how to present knowledge but not what counted as knowledge in the first place. For that, Wikipedia needed additional policies.

Verifiability

Verifiability emerged in August 2003, directly from failures in NPOV enforcement.[82] The physics cranks had exposed a gap: NPOV said to represent significant viewpoints, but cranks claimed their viewpoints were significant. What external standard could distinguish legitimate minority positions from personal theories?

The answer was publication. If a viewpoint had been published by a reliable source, it was verifiable. If it existed only on someone's personal website, it was not. The burden of proof fell on content-adders: if you wanted something in Wikipedia, you needed to cite a source.[82]

The stance was later distilled into the policy's most controversial phrase, added to the official text in 2005: the threshold for inclusion was verifiability, not truth.[82]

This was deliberately provocative. Wikipedia explicitly disclaimed responsibility for determining truth.[82] A claim could be included if a reliable source published it, even if the claim was actually false. A claim could be excluded if no reliable source published it, even if the claim was actually true. Wikipedia would reflect published knowledge, not reality directly.

The physics crank problem was solved. "Your personal website is not a reliable source" became a decisive argument.[82] The crank's theory, however compelling to its author, had not been published in journals. It could not be included. The dispute was over. Not because anyone had proven the theory wrong, but because it failed the verifiability test.

But new questions emerged. What counted as "reliable"? A peer-reviewed journal was clearly reliable. A blog post was clearly not. But what about newspapers with uneven fact-checking, think tanks with ideological agendas, or publishers in developing countries with different standards? The policy provided a principle but left interpretation to case-by-case judgment.[82]

No Original Research

No Original Research, abbreviated NOR, codified in December 2003, drew Wikipedia's last boundary. Wales later acknowledged its origins in a 2004 mailing list post: the phrase "original research," he wrote, "originated primarily as a practical means to deal with physics cranks, of which of course there are a number on the web."[90]

NOR completed the framework. Wikipedia would synthesize existing published knowledge, not create new knowledge. Editors could summarize what sources said. They could not draw novel conclusions, present

unpublished arguments, or introduce information that no reliable source had published.[97]

This meant Wikipedia had a knowledge boundary. Topics without published reliable sources could not have Wikipedia articles.[82] Oral traditions that had never been written down were excluded. Emerging fields that hadn't yet produced peer-reviewed literature were excluded. Communities without publishing infrastructure, disproportionately in the Global South, found their knowledge systematically underrepresented.[98]

The policy's architects knew this was a trade-off. Excluding unpublished knowledge meant excluding valuable knowledge. But the alternative, allowing anyone to add anything they believed true, would destroy Wikipedia's utility as a reference. The fringe theory problem had proven that.

Together, the three policies formed Wikipedia's epistemological architecture. NPOV defined what Wikipedia would say: describe viewpoints, don't assert truth. Verifiability defined how claims would be validated: cite reliable sources. NOR defined what Wikipedia would not do: create new knowledge. The framework gave editors tools to resolve disputes that had previously festered indefinitely.

What Remained Undecided

The standard story of Wikipedia's governance ends with the three policies in place. That story needs revision.

The three core content policies were a remarkable achievement. They enabled millions of editors to collaborate despite fundamental disagreements about truth.[47] They created common ground where none had existed. They solved the fringe theory problem.

But they solved only one kind of problem. The policies addressed content. They did not address power.

And power matters in a commons. Commoners have rights: the right to contribute, to see decisions made transparently, to have a say in how

shared resources are governed. The content policies told contributors how to write articles. They said nothing about who controlled Wikipedia's direction.

Did volunteers have the right to veto commercial partnerships? Could they require transparency about such deals? The policies created a workable system for deciding *what Wikipedia would say*. They created no system for deciding *who controlled Wikipedia*.

Consider what remained undefined after December 2003. Who has ultimate authority over Wikipedia's direction? Wales, informally, but nothing constrained his authority.[45] What happens when the community and the institution disagree? No process existed. The assumption was that they wouldn't disagree, that shared mission would prevent conflict.

Who controls commercialization, partnerships, external relationships? Not addressed. These questions seemed peripheral to building an encyclopedia. What is "consensus" and how is it determined when it breaks down? It remained a principle without a clear mechanism for contested cases.

In December 2003, Wales created the Arbitration Committee to handle intractable disputes. Wikipedia's own institutional history describes ArbCom as an extension of the decision-making power Wales had previously held himself, a revealing characterization.[99] His authority did not dissipate. It institutionalized. ArbCom could ban editors and resolve content disputes. It could not constrain the Wikimedia Foundation, incorporated six months earlier,[100] or determine Wikipedia's commercial relationships, or protect volunteer interests against institutional decisions.

The gap mattered once the Foundation held money, staff, and commercial power. The governance questions these early policies sidestepped became central rather than peripheral, including the question of who really owned what the volunteers had built. Content governance was solved. Institutional governance was left to trust, goodwill, and the assumption that everyone shared the same mission.

The policies' useful vagueness enabled collaboration. The same vagueness also made co-optation possible.

The Argument Over What the Policies Did

To Wales and the mainstream of the community, the policies were a triumph. They solved a real problem: how to scale collaborative knowledge production without gatekeeping experts. They let editors with incompatible worldviews work together productively. They made Wikipedia possible.

The evidence looked compelling. Wikipedia grew from tens of thousands of articles to millions.[47] A 2005 study in Nature found its accuracy broadly comparable to Encyclopedia Britannica's, with somewhat more errors per article, a finding Britannica disputed.[101,102] Edit wars declined after the policies were formalized. Hundreds of language editions adopted the same framework, proving its cross-cultural reach.[47]

"NPOV ensures that we can join the scattered pieces of what we think we know and good faith facilitates the actual practice of fitting them together," Reagle wrote.[45] The policies did not resolve disagreements about truth. They made such resolution unnecessary for collaborative work.

Sanger drew the opposite lesson. To him the policies encoded everything wrong with Wikipedia's direction. His 2004 essay "Why Wikipedia Must Jettison Its Anti-Elitism" named "the root problem: anti-elitism, or lack of respect for expertise." Anti-elitism, as he defined it, meant that "expertise is not accorded any special respect, and snubs and disrespect of expertise is tolerated."[74]

Verifiability, in this view, privileged publication over knowledge. A tenured physicist could not override a published crank theory on the basis of expertise.[82] The policy's refusal to claim truth struck critics as cowardice. Wikipedia claimed no responsibility for accuracy while becoming the world's de facto reference.[4] When errors persisted in specialized articles, no mechanism let experts correct them authoritatively.[74]

Critical scholars pushed further still. The policies looked neutral, they argued, but encoded power relations invisible to their architects. "Wikipedia's infrastructure introduces new and less visible sources of gender disparity," Heather Ford wrote with Judy Wajcman.[98] The same could be said for geographic disparity, linguistic disparity, and disparities of all kinds.

NOR excluded oral traditions. Verifiability excluded unpublished knowledge.[82] NPOV's requirement for "significant" viewpoints meant that marginalized perspectives, significant to their communities but not to mainstream publishers, could be dismissed as fringe. From this angle the fringe theorists were a distraction. The real exclusions were not crackpots with personal websites. They were entire knowledge traditions that had never been published in forms Wikipedia could recognize.[98]

The debate never reconciled, because the policies' constructive vagueness cut both ways. It resolved content disputes and deferred the questions of power.

Five Pillars

On May 4, 2005, a Wikipedia editor with the username Neutrality created a new page: "Wikipedia:Five pillars." The page distilled five claims. Wikipedia was an encyclopedia, written from a neutral point of view, free to reuse, governed by civility among editors, and not bound by firm rules.[103]

There was no community discussion. No vote. No formal process. One editor created a page that summarized existing norms, and the community accepted it. Within months, the Five Pillars became Wikipedia's de facto constitution, cited in countless disputes and orientation materials.[103]

This was the triumph and the trap of Wikipedia as a commons without formal governance. The Five Pillars arose through the most democratic possible process: emergence without authority. One editor summarized

norms that already existed. The community recognized them as accurate. Norms could be created by consensus.

But they could also be abandoned by consensus, and consensus itself was never institutionalized. A real constitution has enforcement mechanisms. The Five Pillars had only agreement. When commercial interests began to threaten commons principles, the Pillars would offer no legal recourse, no institutional protection.

This pattern, emergent governance through individual initiative accepted through presumed consensus, characterized Wikipedia's institutional development. It worked when participants shared goals. It would fail when they did not.

Part II

The Licensing Gamble

CHAPTER FIVE

The Copyleft Promise

> GNU is not in the public domain. Everyone will be permitted to modify and redistribute GNU, but no distributor will be allowed to restrict its further redistribution.
>
> — Richard Stallman, GNU Manifesto, 1985

One day in the early 1980s, Richard Stallman walked into an equipment room at the MIT Artificial Intelligence Laboratory and found the PDP-10 sitting dark and silent.

The machine had been the heart of a community. For more than a decade, some of the world's best programmers had gathered around it, sharing code freely, building on each other's work, treating software as a communal resource. They called themselves hackers, a term that meant something different then. It meant someone who loved to program, who sought elegant solutions, who believed that information wanted to be free.

The indicator lights were out. The fans were silent.

Stallman stood alone in the equipment room and began to cry.

"Seeing the machine there, dead, with nobody left to fix it," he later recalled, "it all drove home how completely my community had been destroyed."[21]

[1]Stallman's account appears in Sam Williams's biography. He recalled: "I started crying right there in the equipment room."

The destruction hadn't been sudden. It had come gradually, as commercial interests colonized what had once been shared space. Two companies had hired away most of the lab's programmers, and the code that had been freely shared became proprietary. The improvements that had benefited everyone now belonged to corporations.[2]

By the time the PDP-10 went dark, the war was already lost. The hacker community that had formed around it was gone.

The destruction of his community left Stallman with a question that would define his life's work. His answer would shape the legal framework for free knowledge for the next forty years. It would become the foundation of Wikipedia's promise to its contributors.

Lab Culture

Stallman joined the MIT Artificial Intelligence Laboratory in 1971, at age eighteen. The lab embodied a distinct philosophy about software: code was shared freely, improvements were collective, and no one owned what everyone built.[2]

This was not idealism. It was practical. When one programmer improved a piece of code, everyone benefited. When someone found a bug, the fix went into the common codebase. The machine and the humans around it formed a collaborative whole, each contributing what they could, each using what they needed.

"For more than ten years, many of the world's best programmers worked at the Artificial Intelligence Lab for far less money than they could have had anywhere else," Stallman later wrote in the GNU Manifesto. "They got many kinds of nonmonetary rewards: fame and appreciation, for example. And creativity is also fun, a reward in itself."[3]

The lab had no passwords initially. When MIT's Laboratory for Computer Science installed password controls in 1977, Stallman cracked them and offered fellow users a way to opt out.[2] Security was antithetical to sharing.

What governed this shared resource was a set of norms and expectations, what we might call a commons. No formal licensing existed yet, but the principle was clear: those who benefited from collective work had an obligation to return their improvements.[2]

This culture had limits. Two companies formed to commercialize the Lisp Machine, a specialized computer the lab had developed. The communal model collided with commercial interests.[2]

Symbolics War

Lisp Machines, Inc. (LMI) emerged from the AI Lab in late 1979, and Symbolics followed in early 1980, each seeking to profit from technology the community had built together.[2] This was the first enclosure: the appropriation of commons resources by commercial entities that refused to return their improvements.

Initially, a "gentleman's agreement" allowed Stallman to review Symbolics' improvements. He could incorporate them into the common codebase, ensuring the community still benefited.[2]

On March 16, 1982, Stallman's twenty-ninth birthday, Symbolics ended that agreement. They would no longer share their improvements. If the lab wanted the latest features, they would have to abandon the community approach entirely.[2]

Stallman interpreted this as an act of war. "I was going to punish Symbolics if it was the last thing I did," he said. For two years, he cloned every feature Symbolics developed, working alone against an entire company. His framing was martial: "The AI Lab was a neutral country, like Belgium in World War I. If Germany invades Belgium, Belgium declares war on Germany and sides with Britain and France."[2]

Yet this was a holding action, not a victory. The lab's programmers were leaving. The community was fragmenting. No single person, however dedicated, could replace what collective effort had built.

Some of Stallman's colleagues saw things differently. They viewed Symbolics not as an invader but as a necessary evolution, bringing hacker principles into the commercial marketplace. Many hackers, as Stallman's biographer Sam Williams put it, saw him as "a troubling anachronism," clinging to a model that couldn't survive contact with economic reality.[2]

The social isolation was real. Richard Greenblatt, one of the lab's central hackers and a founder of LMI, had begun a custom of inviting everyone along to group dinners. "I was no longer getting invited to go to Chinatown," Stallman recalled. "The custom started by Greenblatt was that if you went out to dinner, you went around or sent a message asking anybody at the lab if they also wanted to go. Sometime around 1980-1981, I stopped getting asked."[2]

This was not just a philosophical dispute. It was personal.

GNU Manifesto

On September 27, 1983, Stallman announced his response on a Usenet newsgroup, an early online discussion forum: he would build an entire operating system that could never be taken from its users.[2]

He called it GNU, a recursive acronym for "GNU's Not Unix." The project would recreate the functionality of Unix, the dominant operating system, but with one decisive difference: GNU would be free forever. Not free as in price, but free as in freedom. Users would have the right to run it, study it, modify it, and share it.[3]

On January 5, 1984, Stallman quit his job at MIT to work full-time on GNU.[2] He wanted to avoid any claim that MIT owned his code. In March 1985, the GNU Manifesto appeared in Dr. Dobb's Journal, laying out both the practical plan and the philosophical foundation.[3]

The manifesto articulated what would become copyleft. Proprietary modifications would not be allowed. Every version of GNU would remain free. Stallman had learned from Symbolics that informal sharing

agreements could be revoked. Only legal architecture could ensure permanence.[3]

This was different from simply giving software away. Public domain software could be taken, modified, and locked up, exactly what had happened at the AI Lab. Stallman's innovation was to use copyright law against itself: the software would be copyrighted, but the license would require that anyone who modified and distributed it must preserve the same freedoms.[3]

The license was a weapon. It was designed to punish exactly what Symbolics had done: taking shared work and refusing to share the improvements.[3]

Four Freedoms

Stallman codified what "free software" meant through four freedoms. Freedom 0 was the freedom to run the program as you wish, for any purpose. Freedom 1 was the freedom to study how the program works and change it, which requires access to source code. Freedom 2 was the freedom to redistribute copies so you can help others. Freedom 3 was the freedom to distribute copies of your modified versions, giving the whole community a chance to benefit. The numbering was significant: Freedom 0 was added around 1990 when Stallman realized the freedom to simply use the program needed explicit protection.[104]

These freedoms were not merely desirable. They were essential. Without them, users were subject to the software's owners. With them, users controlled their own computing.

License Evolution

Copyleft's legal framework evolved through hard lessons.

In 1985, Stallman published the first copyleft license, specific to GNU Emacs, the text editor he had built. This became necessary after the Gosling Emacs incident: James Gosling had written his own version of

Emacs in 1981, initially allowing free distribution. Stallman incorporated some of this code into GNU Emacs. When Gosling sold the rights to UniPress, the company forced Stallman to remove the code. The incident confirmed what Symbolics had taught: only a license could make sharing permanent.[2]

The GNU General Public License (GPL) version 1 appeared in February 1989, the first program-independent copyleft license. Any project could adopt it, making the license itself a reusable component.[105]

GPL version 2 followed in June 1991, adding the "Liberty or Death" clause. If you cannot comply with all GPL terms because of patent claims or other legal restrictions, you cannot distribute the software at all. This prevented GPL circumvention through external legal constraints.[106]

Each step responded to attempts at enclosure. Each built higher walls around the commons.

Beyond Software

By the late 1990s, Stallman recognized that software freedom was incomplete without documentation freedom:

"The biggest deficiency in free operating systems is not in the software—it is the lack of good free manuals that we can include in these systems. Many of our most important programs do not come with full manuals. Documentation is an essential part of any software package; when an important free software package does not come with a free manual, that is a major gap."[107]

The problem was that good manuals were being published with restrictive terms. O'Reilly Associates, for example, produced excellent documentation but prohibited copying and modification. These manuals were "nonfree" and thus "excluded from the Free World."[107]

The GNU Free Documentation License (GFDL), released in March 2000, extended copyleft to documentation. It required attribution, preserved modification history, and ensured derivatives remained free. It also

included provisions for "invariant sections," portions that could not be altered, designed to protect philosophical statements from commercial publishers who might want to strip the ideology from the documentation.[108,5]

License Assumptions

The GFDL is where the standard story of copyleft needs revision.

The GFDL was designed to prevent a specific form of enclosure: commercial entities taking free documentation, improving it, and refusing to share their improvements. The same logic carried from software to text.

The license assumed that derivatives would be made by humans, distributed as human-readable works, and used by other humans. Its requirements followed from that assumption. Bundling the license text with every copy, preserving the history, crediting the authors: each made sense because a "derivative work" meant a recognizable modification, a revised manual or an updated textbook or a translation. "Distribution" meant handing copies to other people. Attribution could be verified by reading the work itself.[5]

The license could not anticipate a form of use that would transform content into statistical weights. No one could have. The outputs echo the training data, sometimes nearly verbatim, produced at industrial scale without human review.[9,109]

When AI companies train on Wikipedia, they create something that may not legally qualify as a "derivative work." The model weights are not a modified encyclopedia. The outputs are not edited articles. Copyleft propagates through derivatives, and that mechanism may have no purchase on this new form of extraction.[17,69]

But this is not a failure of Stallman's vision. The legal arguments for enforcement exist. Making them requires an institution willing to litigate.

The Copyleft Argument

From Stallman's perspective, and that of the Free Software Foundation he founded, copyleft was essential protection. The alternative was permissive licenses like BSD and MIT, which allow proprietary derivatives.[110] Under those terms free software became a gift to commercial developers. They could take it, improve it, and lock it up, and the community's work would benefit corporations without any reciprocity.

"The fundamental act of friendship among programmers is the sharing of programs; marketing arrangements now typically used essentially forbid programmers to treat others as friends," Stallman wrote.[3]

Copyleft made freedom propagate. Anyone could use, modify, and distribute the work, but only if they preserved the same freedoms for others. The mechanism was often called "viral": freedom spread through every derivative. The evidence supported the view. Linux, the kernel that completed the GNU operating system, was released under the GPL. The combination became the foundation of the modern internet. Servers running free software powered the web, and commercial entities contributed improvements back because the license required it.[2,110]

Not everyone embraced this maximalism. Critics within the open source community argued that copyleft's "viral" nature created unnecessary complexity.[50] Businesses feared inadvertently contaminating proprietary code. Simpler licenses such as MIT, BSD, and Apache achieved broader adoption without the compliance burden.

The market appeared to validate the complaint. Copyleft licenses, once used by a majority of open source projects, fell to roughly a fifth of them by the mid-2020s. Permissive licenses gained ground.[110] Cloud computing let companies use copyleft software without triggering distribution requirements, since running software as a service was not legally "distribution."[111] By this account Stallman's approach was ideologically pure but practically costly. The goal should be maximum adoption, not maximum protection of ideological freedoms.

Documentation raised a separate objection. The GFDL had been designed for software manuals, not encyclopedias.[45] Its requirements, the invariant sections and cover texts and pages of license text with every copy, made sense for printed manuals distributed by publishers. They made less sense for wikis, where content was fluid, authorship was distributed, and reuse might mean excerpting a single paragraph. Wikitravel's founders put it bluntly when they chose CC-BY-SA over the GFDL in 2003: requiring people to pass out pages of legalese with every small document defeated the purpose of a free license.[112] But in 2001, when Wikipedia launched, the GFDL was the only credible copyleft license for documentation. Creative Commons, the organization behind CC-BY-SA, did not exist yet.[113]

The debate set maximum protection against maximum adoption, ideological purity against pragmatic reach. Stallman would rather have fewer users than enable enclosure. The pragmatists would rather have more people using free software, even if some built proprietary derivatives. Both camps assumed enforcement would happen. Neither asked whether it actually would. That question passed, almost unnoticed, into Wikipedia's license choice.

Wikipedia Enters

In January 2001, after email exchanges with Richard Stallman himself, Jimmy Wales committed Nupedia, his expert-written encyclopedia project, and then Wikipedia, to the GFDL.[36]

This was not a casual choice. The GNU Project noted: "Just as we were starting a project, GNUpedia, to develop a free encyclopedia, the Nupedia encyclopedia project adopted the GNU Free Documentation License and thus became a free commercial project [sic]. So we decided to merge GNUpedia project into Nupedia."[1][42]

[2]The phrase "free commercial project" appears verbatim on the GNU Project's page; in context it almost certainly means a free-content project.

Wikipedia entered Stallman's ideological framework. The project accepted his legal architecture, his philosophical commitments, and his assumptions about how protection would work.

Contributors who edited Wikipedia understood the deal: their work would remain free forever. Anyone could use it, modify it, share it, but derivatives would have to preserve the same freedoms. Commercial publishers couldn't lock up volunteer labor. The copyleft mechanism would ensure that what was shared stayed shared.[45]

This was the promise.

But the promise rested on assumptions no one questioned, and those assumptions would be tested from two directions at once. AI training would turn content into model weights whose status as derivative works was contested.[17] And the organization stewarding the license would become financially entangled with the very companies whose compliance was questionable. It would be a guardian whose cultural networks overlapped with the extractors, one that viewed them not as adversaries but as colleagues.

Stallman had stood before the dead PDP-10, watching a community's destruction. His response was to build legal defenses against future enclosures. Those defenses worked for decades, protecting free software and free documentation from commercial appropriation. The institution that inherited them would be tested by enclosures Stallman never imagined.

In 2001, Wikipedia made a promise to its contributors: your work will remain free forever. That promise was encoded in the GFDL, backed by Stallman's legal architecture, inherited from the grief of a programmer who watched his community die.

Tens of thousands of editors would contribute under that promise. Millions of articles would be created. The largest reference work ever assembled would emerge from collective volunteer labor, protected by copyleft.[93]

Twenty-three years later, that work would train systems worth hundreds of billions of dollars. The editors would receive no attribution. The derivatives would share nothing alike. The promise would become theoretical.

The tragic flaw lay not in the protection but in who inherited it. Stallman had built the license as a weapon. Wikipedia inherited it and never once raised it.

CHAPTER SIX

The GFDL Years

> For such small documents, it just doesn't make sense to require people to pass out another 10 pages of legalese text.
> — Wikitravel, "Why Wikitravel isn't GFDL"

Imagine a teacher in 2006 who wants to give her students a one-page handout on photosynthesis. Wikipedia has an excellent article on the topic: clear, accurate, well-illustrated. She can legally copy it. The license says so.

She opens a word processor, pastes the text, and starts formatting. Then she reads the license requirements.

The GNU Free Documentation License required that any copy include the full license text. Several pages of legal language.[115] For a one-page handout, she would need to attach the entire license.

The license also required listing the "principal authors." But the Wikipedia article had been edited by hundreds of people. Whose names should appear? The license required preserving the complete revision history.[115] For a classroom handout?

She closes the word processor and types the handout from scratch, summarizing the article in her own words. The license that promised "free knowledge" has made the article easier to rewrite than to copy.

Real reusers hit this wall constantly. It had a name: the flier problem.[116] And it was only the beginning.

January 2001

When Wikipedia launched, there was no real alternative. Creative Commons would not release its first licenses until December 2002, nearly two years later.[117] For a project committed to copyleft, where every reuse must carry forward the same freedoms, the GFDL was the only door.[118]

Jimmy Wales had corresponded with Richard Stallman in early January 2001 about licensing.[36] Wales was not a free software ideologue. He was an entrepreneur whose web portal, Bomis, ran on advertising revenue.[44] But he understood that a free encyclopedia needed a free license, and Stallman was the expert.

As one anonymous Slashdot commenter would later quip, recalling those early days: "Having Richard Stallman brow beat Jimmy Wales certainly didn't hurt either."[119] It was hearsay, offered eight years after the fact, and even the commenter doubted it changed much.

The choice was pragmatic, not ideological. Bomis had shown Wales that advertising could fund free content, and the GFDL's copyleft requirements were a constraint he accepted rather than a creed he embraced. The tension between free-content ideology and commercial logic was built into the project from the start, and it outlasted any one person.

GFDL Requirements

The GFDL imposed specific obligations that made sense for software manuals. Every copy had to include the complete GFDL, several printed pages. The complete modification history had to be preserved, including dates and authors. At least five authors had to be listed on the title page, or all of them if there were fewer than five. The source format (what the license called a "transparent copy") had to be made available, not just the rendered output. If the original specified cover texts (short required blurbs

printed on the front or back cover), these had to be preserved. Certain sections designated as "invariant" (fixed passages the author marked off-limits) could not be modified or removed.[115]

Wikipedia prohibited invariant sections and cover texts, which avoided some complexity.[116] But the other requirements applied fully.

For a printed software manual, these requirements were manageable. The manual already had a title page. The revision history already existed. Those pages of license text were negligible in a 500-page technical reference.

For a wiki, they were awkward. Who were the "principal authors" of an article edited by hundreds of people? How could a revision history be preserved when every edit created a new version?[120] What did "transparent copy" mean when the source was a wiki page that changed constantly?

Practical Problems

By the mid-2000s, the GFDL's awkwardness for Wikipedia had become a running joke in the community.[116]

The flier problem generalized. Any physical reproduction required the full license text.[115] Teachers couldn't make classroom handouts, and community organizations couldn't print educational materials.[112]

The audio and video problem was worse. Creating a video based on Wikipedia content meant what, exactly? Reading the full license aloud? Displaying it as on-screen text? The license was written for documents. It had no practical application to multimedia.

The image problem. The GFDL was written for documentation and text, not images.[5] A photograph released under GFDL technically required the full license text with every use.[115] A 10-kilobyte image would need a 50-kilobyte text file.

The many-authors problem. It became absurd at scale. By the mid-2000s, popular articles like "United States" had accumulated thousands

of contributors. Listing the "five principal authors" required determining who had contributed most, a computation nobody performed.[115] The result was universal non-compliance.

The legal uncertainty problem. It was perhaps the most troubling. As longtime Wikipedia administrator David Gerard later put it: "Literally no one understands how to reuse GFDL content safely, including the FSF."[121]

The Free Software Foundation had created the license.[5] Wikipedia was its largest user. And the FSF would not clarify how the license applied.[121]

Defections

Other projects looked at Wikipedia's struggles and chose differently.

In 2003, Wikitravel launched under CC-BY-SA (Creative Commons Attribution-ShareAlike), a simpler license, explicitly rejecting the GFDL. Its founders, Evan Prodromou and Michele Ann Jenkins, judged the software-manual-oriented license unworkable for a travel guide.[122] The project's licensing page, titled "Why Wikitravel isn't GFDL," made the objection concrete: a traveler should be able to print a page or two and hand it to someone without an appendix of legalese.[112]

In March 2006, Debian, the respected free software distribution, voted on whether GFDL documents could be included in its main repository. The result: documents with invariant sections were "not suitable for Debian main," though GFDL works without them, like Wikipedia's, were deemed free.[123] Even so, the verdict was a legitimacy blow. The free software community itself had judged parts of the FSF's own documentation license non-free.

In 2007, FLOSS Manuals abandoned the GFDL for the GPL (the license Stallman wrote for software), citing "difficulties in implementing the GFDL."[124]

Jonathan Corbet of the free-software news site LWN.net summarized the emerging consensus: "The GFDL has come to be seen by many as

more of a tool for the propagation of FSF propaganda than a license for truly free documentation. Much of the community avoids this license. Some groups, such as the Debian Project, see it as non-free."[116]

The defections left Wikipedia isolated. Content under the GFDL could not be legally combined with content under CC-BY-SA. Both were copyleft. Both required share-alike, meaning any derivative had to carry the same license forward. But their terms differed in ways that made combination impossible. Wikitravel, Citizendium (Sanger's rival project), and WikiEducator, an educational-materials wiki: the newer wikis all chose CC-BY-SA.[122,125] The free content world was fragmenting along licensing lines, and Wikipedia's GFDL commitment, made when no alternative existed, had become a barrier to collaboration.

Friction as Protection

For years the GFDL's friction looked like pure defect, the fault of a clumsy license finally escaped in 2009. It was also the protection.

The same mechanisms that made the GFDL cumbersome for legitimate reuse made it cumbersome for extraction. The requirements that annoyed teachers also deterred scrapers.

Consider what the GFDL's "inconvenient" requirements actually did. Full license text with every copy created overhead that discouraged bulk extraction. Complete history preservation built an audit trail and complicated the laundering of content. Principal-author attribution kept human creators visible. Transparent-copy availability kept the source accessible. Share-alike for every derivative meant anything built on Wikipedia inherited these same burdens.[115]

A teacher could not make her flier without attaching the full license. Neither could a commercial entity harvest Wikipedia's entire corpus without inheriting that license, for every article. The complaints measured convenience for legitimate small-scale reuse. They did not measure protection against industrial-scale extraction.

The same trade shaped the later move to CC-BY-SA. Its flexible attribution, where a link suffices and no author list is required, made legitimate reuse easier.[16] It made extraction easier too.

Escape Hatch

Any escape from the GFDL rested on a clause few had read closely. Wikipedia's copyright notice licensed content under "GFDL 1.2 or any later version published by the Free Software Foundation."[115]

This routine provision, included without much thought, gave the FSF power to create new versions that Wikipedia could automatically adopt.[116] As Corbet noted:

"The other thing to be aware of is just how much power the 'or any later version' text puts into the hands of the FSF. The license promises that later versions will be 'similar in spirit,' but the GPLv3 debate made it clear that similarity of spirit is in the eye of the beholder."[116]

The FSF had built what looked like permanent walls around the shared resource. But it had left a key in the lock, a power to change the license if times required. That power made sense for software. For a collective work, it created a vulnerability. If the institution guarding the commons changed its mind, or became entangled with extractive interests, the walls could be remade from within.

Whose Consent?

When the push to escape the GFDL gathered force, it split the community into camps that never fully reconciled.

For the Wikimedia Foundation and most of the community, the case was simply practical. The license was genuinely unsuitable for Wikipedia's use case. The friction was real, documented in thousands of frustrated emails and forum posts. CC-BY-SA preserved copyleft while enabling practical reuse.[16] "When I started Wikipedia, Creative Commons did not exist," Wales explained. "The Free Documentation License was the first

license that demonstrated well how the principles of the free software movement could be applied to other kinds of works." But times had changed. CC-BY-SA was "a more generic license that meets the needs of Wikipedia today."[118]

A minority asked whether simplification was worth the cost. When contributors licensed their work, no one understood the "or any later version" clause to authorize changes this fundamental. Contributors who chose the GFDL in 2001 or 2005 had not consented to CC-BY-SA. They had consented to trust the FSF with future versions. One contributor, Chris Frey, wrote to the FSF calling the forthcoming migration a "breach of trust."[126]

Slashdot commenters worried: "It is slightly chilling for anyone using another FSF license. You can omit the 'or later versions' license and have the possibility that the later versions of other FSF licenses will be incompatible with your version. Or you can include it and have the possibility that the FSF will decide to grant an exemption for a specific large organisation and allow them to relicense your work."[127] The objection was not that CC-BY-SA was bad. It was that contributor expectations could be overridden when institutional interests aligned.

Richard Stallman saw no betrayal. The FSF had never promised its licenses would stand still, he argued, only that changes would keep to the spirit of the original and uphold the purposes it was written for.[126]

The tension between stability and evolution was never resolved. But a precedent was set: when institutions agree, contributor expectations can be adjusted.

The Exit Permit

The Foundation moved to make the escape real. On December 1, 2007, the Wikimedia Foundation Board voted 5 to 0, with one trustee not voting, to ask the Free Software Foundation to modify the GFDL so Wikipedia could migrate to CC-BY-SA.[128]

A year of private negotiations followed, involving the Foundation, Creative Commons, and the Software Freedom Law Center.[126] On November 3, 2008, the FSF released GFDL 1.3, a tailored later version that opened a time-limited window for Wikipedia to relicense.[116] The exit permit had been granted. What the community would actually do with it still lay ahead.

What Was Traded

The teacher with her photosynthesis handout would soon be able to share Wikipedia content with a link and nothing more. No pages of license text.[16] This was real relief from a real burden.

The flier problem, the defections to CC-BY-SA, the running joke of a license no one could safely obey: every complaint pushed the same way. The burden had become unbearable, and pressure to escape the GFDL had been building for years.[116] The exit permit was now waiting.

By late 2008, the legal obstacle had become a political question. Years of friction had made the license's removal feel inevitable.

CHAPTER SEVEN

The First Loosening

> How could a small part of the community vote for a license change affecting all contributing right owners?
>
> — Herkko Hietanen, WIPO Magazine, 2009

On November 30, 2007, at a Creative Commons party in San Francisco, Jimmy Wales made an announcement. Creative Commons, Wikimedia, and the Free Software Foundation had agreed to make Wikipedia's license compatible with Creative Commons. This, Wales declared, was the "party to celebrate the liberation of Wikipedia." The Wikimedia Foundation's board formalized the request the next day.[128]

The crowd cheered. These were people who understood licensing. They had felt the friction of GFDL. They wanted Wikipedia to join the broader Creative Commons world, and they saw the migration as progress.

Most Wikipedia contributors were not at this party. They were scattered across the world, connected only by their shared project: the retired teacher in Ohio, the physics student in India, the hobbyist in rural England. Decisions about the global commons were made by a local elite. The millions who had contributed over the previous seven years were not present and would not be asked. The party celebrated liberation. The question was liberation for whom.

Secret Negotiations

The path from the GFDL to CC-BY-SA required a year of private talks between four organizations: the Free Software Foundation, the Wikimedia Foundation, Creative Commons, and the Software Freedom Law Center.[126]

Richard Stallman explained the secrecy: "We held these discussions in private for a reason I hope you will approve of: to keep the option limited in applicability and avoid the possibility of wholesale relicensing of other GFDL-covered material."[126]

The negotiations were tailored to Wikipedia's situation. The FSF was not opening a general exit from GFDL. It was creating a one-time window for a specific type of project.[126,116]

The participants had overlapping interests. The FSF wanted to support Wikipedia without undermining GFDL generally. The Wikimedia Foundation wanted to escape GFDL's friction. Creative Commons wanted to unify free content under its licenses. The Software Freedom Law Center handled the legal machinery.[126,15,125]

The result was GFDL 1.3, released November 3, 2008, with Section 11, the escape clause.[116]

Exit Permit

Section 11 of GFDL 1.3 created the concept of "Massive Multiauthor Collaboration Sites," wikis in all but name. It gave them a time-limited option to relicense to CC-BY-SA.[116]

The requirements were specific. Content had to originate on an MMC Site, not imported from external GFDL sources after November 1, 2008. It had to be licensed under GFDL with the "or any later version" clause and carry no cover texts or invariant sections. Migration had to occur before August 1, 2009.[116]

Jonathan Corbet at LWN.net called it "a legal hack": "Legal codes, like other kinds of code, have a certain tendency to pick up cruft as they are patched over time. In this case, the FSF has added a special, time-limited

hack which lets Wikipedia make a graceful exit from the GFDL license regime."[116]

The mechanism was elegant if you accepted its premises. The "or any later version" clause, which contributors had included routinely since 2001, gave the FSF authority to create new versions. GFDL 1.3 used that authority to enable migration to CC-BY-SA. The chain was complete: the original license permitted later versions, GFDL 1.3 was a later version, and GFDL 1.3 permitted CC-BY-SA.[116]

The whole migration rode on a clause most contributors had never examined. Those who licensed under GFDL in 2001 had trusted the FSF with future versions of GFDL. They had not consented to migration to a different license family. The chain of authority was legally valid, probably. The consent behind it was attenuated.[116]

The Vote

The community vote on the migration ran from April 12 to May 3, 2009, twenty-one days. Eligibility required 25 or more edits to any Wikimedia project before March 15, 2009.[129]

The rule excluded large parts of the community. Casual contributors with fewer than 25 edits could not vote, though their contributions were counted in the work being relicensed. Anonymous IP editors, who wrote a large share of Wikipedia, had no voice. Inactive contributors who had drifted away by 2009 were unlikely to hear about the vote. Contributors who had died in the eight years since 2001 could not participate.

Notification came through CentralNotice, which displayed banners to logged-in users. Those who were not logged in, or were not editing during the voting period, might never learn the vote was happening. Many eligible voters never did.

The ballot asked voters to approve three changes: dual-licensing under CC-BY-SA 3.0 and GFDL, making CC-BY-SA the primary license going forward, and adopting a foundation-wide attribution policy. Software in

the Public Interest (SPI), a nonprofit that provides infrastructure for open source projects, managed the voting.[129]

The vote produced the results shown in Table 7.1.

Metric	Count
Total ballots submitted	18,692
Valid ballots (after review)	17,462
Yes votes	13,242 (75.8%)
No votes	1,829 (10.5%)
No opinion	2,391 (13.7%)

Table 7.1. 2009 CC-BY-SA migration vote results

Rejected ballots included 930 multiple votes from single accounts, 256 votes from the same person using different wiki accounts, and 44 other violations. Excluding abstentions, 87.9 percent of those who expressed an opinion supported migration.[129] The result was celebrated as a decisive democratic mandate.

Herkko Hietanen, writing in WIPO Magazine, the publication of the World Intellectual Property Organization, asked the question the celebration obscured. How could a fraction of the community bind all contributors?[15]

By 2009, Wikipedia had millions of historical contributors and hundreds of millions of edits across all projects. The English edition alone held roughly 10 million registered accounts, plus countless anonymous IP contributions.[130] The vote produced roughly 17,500 valid ballots. About 13,000 people voted yes.[129] Thirteen thousand people relicensed the work of millions.

WIPO's analysis raised a deeper concern:

"It is a matter of legal debate as to whether clauses covering future possibilities are valid in the case of licensors who were unaware of exploita-

tion or licensing options not yet invented at the time of initial distribution. Most European countries have laws nullifying such agreements."[15]

The "or any later version" clause enabled the chain. But contributors in 2001 had not anticipated, and could not have anticipated, that it would be used to migrate to a different license family entirely.

What Changed

The migration was not merely a format change. CC-BY-SA carried substantively lighter requirements than GFDL. Under GFDL, every copy required the full license text, roughly five pages. CC-BY-SA dropped that. GFDL required preserving the complete modification history, and CC-BY-SA did not. GFDL required listing at least five principal authors, while CC-BY-SA let a link to the source suffice. GFDL required a transparent copy in source format, which CC-BY-SA dropped entirely.[15,131]

The same requirements that made GFDL inconvenient for a classroom handout also made it inconvenient for anyone who wanted to extract Wikipedia at scale. The simpler license was also a weaker one. The friction removed for legitimate reuse was removed for industrial extraction too.

Every step assumed that enforcement would follow. The contributors who licensed under GFDL in 2001 bet that the FSF would not fundamentally change the terms. The Foundation bet that the commons could migrate to a simpler license without losing protection. Both bets assumed some organization would actually use the legal tools to keep derivatives free.

The Legitimacy Debate

The Foundation and most active editors held that the vote was legitimate and necessary. The practical case was the one the previous chapter assembled: friction that was real, documented, and widely complained about, and a successor license that preserved copyleft while enabling practical reuse.[116] The community had voted, and 88 percent of those expressing an opinion approved.[132] This was democracy in action.

Michael Snow, the Foundation's board chair, stressed continuity in announcing the result: "The volunteers who work on Wikimedia projects have very strongly supported making their contributions available under the Creative Commons Attribution/Share-Alike License (CC-BY-SA) in addition to the GNU Free Documentation License (GFDL)."[132] Erik Moeller, then deputy director, called it "a big day for free culture."[133]

Critics countered that the vote bound people who never cast one, and that consenting to later versions of GFDL was not the same as consenting to a different license family. Chris Frey had already named the stakes in his letter to the FSF: a "breach of trust."[126] On the technology news site Slashdot, others saw troubling precedent. Anyone using an FSF license with the "or later" clause had just learned that their work could be relicensed without their participation.[127] The turnout gap, critics argued, made the result morally thin even if procedurally valid.[15]

The absence of a legal challenge did not settle the question.

Stallman defended the migration: "We have never asserted that we will not change our licenses, or that we will never make changes like this one. Rather, our commitment is that our changes to a license will stick to the spirit of that license, and will uphold the purposes for which we wrote it."[126] Both GFDL and CC-BY-SA were copyleft licenses. Both required share-alike. The essential freedoms survived. The FSF held authority over future versions precisely to enable this kind of evolution. The migration was legitimate, in this view, because the mechanism was legitimate and the destination preserved copyleft's spirit.

Beneath the debate lay one unresolved tension: organizational authority against individual consent. The pragmatists won in 2009. The vote happened, the migration proceeded, and Wikipedia joined the Creative Commons fold. No court ruled on whether "or any later version" authorized this kind of relicensing. The precedent stood. When organizations agree, license terms can evolve beyond what individual creators foresaw, through processes most of those creators never joined.

The Window Closes

On May 21, 2009, the Wikimedia Foundation board passed the Licensing Update Approval Resolution.[134] "Updating our license terms will support Wikimedia's charitable mission," Snow said.[132]

On June 15, 2009, the migration took effect.[134] CC-BY-SA 3.0 became the primary license. New contributions from Wikipedia editors were dual-licensed under CC-BY-SA and GFDL by default. Articles that incorporated externally sourced CC-BY-SA-only content could no longer carry the GFDL at all.[15]

On August 1, 2009, the GFDL 1.3 migration window closed permanently. Any wiki that had not migrated lost the chance forever.[116]

Wales had called it the liberation of Wikipedia. In one sense it was. The flier problem was solved. Wikipedia could now share material more easily with other Creative Commons projects.[125]

But the celebration was for those who understood licensing, joined the process, and believed the vote was legitimate. The millions who had contributed under GFDL were not at the party. They were not asked. They were relicensed anyway. The 2009 switch was the first loosening.

The shape of it would recur. A shared resource had legal protections. Those protections proved inconvenient. Institutions negotiated to reduce them. Consent came through a mechanism most contributors never joined. The result was weaker terms framed as progress.

The precedent of institutional override made future enforcement harder to demand. Organizational interests prevailed, and no one sued, because the community had already acquiesced.

The switch was complete. A few thousand voters had bound millions of contributors, and the legitimacy gap that opened never closed. It waited, unexamined, for the day a new set of companies would arrive to take what the volunteers had never knowingly given.

CHAPTER EIGHT

The Gaps in Practice

> Wikipedia neither condemns, nor promotes, such efforts at enforcing the licenses, and is not responsible for any consequences.
>
> — Wikimedia Meta-Wiki

In September 2005, a friend contacted John Seigenthaler with disturbing news. Something was on the internet about him. Something bad.

Seigenthaler was 78 years old. He had been the founding editorial director of USA Today and an assistant to Attorney General Robert Kennedy. For half a century he had been a respected figure in American journalism. He found his biography on Wikipedia, a website he barely knew.[4]

The text accused him of involvement in the assassinations of both Kennedy brothers.

"John Seigenthaler Sr. was the assistant to Attorney General Robert Kennedy in the early 1960's. For a brief time, he was thought to have been directly involved in the Kennedy assassinations of both John, and his brother, Bobby. Nothing was ever proven."[135]

The entry also claimed, just as falsely, that Seigenthaler had lived in the Soviet Union from 1971 to 1984. For 132 days, this libel had been

available to anyone who searched his name. A false accusation of political murder, presented as encyclopedic fact.[135,4]

Seigenthaler investigated. The vandalism had been added on May 26, 2005, by an anonymous IP address. No one had noticed. No one had checked. For more than four months, anyone who looked up his biography learned that a living American had been suspected of killing two Kennedys.[4]

"When I was a child, my mother lectured me on the evils of 'gossip,'" Seigenthaler wrote in USA Today on November 29, 2005. "She held a feather pillow and said, 'If I tear this open, the feathers will fly to the four winds, and I could never get them back in the pillow.'" The internet, he had learned, was a torn pillow. His verdict: "a flawed and irresponsible research tool."[135]

Wikipedia treated the episode as a content-quality failure, and it built machinery to prevent a repeat. Within months it restricted anonymous article creation, tightened the rules on biographies, and gave editors new tools to erase libel.[47] The gap it cared about, it moved to close. The same years exposed a second gap, deeper and quieter. No one was enforcing the license meant to keep the content free.[136] Wikipedia never built the machinery to close that one. Closing one gap while ignoring the other set the pattern for everything that followed.

The Gaps Wikipedia Closed

Two weeks after Seigenthaler's op-ed, Nature published a study that seemed to contradict everything he said.

The journal had commissioned a comparison of Wikipedia and Encyclopedia Britannica. Experts reviewed 42 article pairs, blinded to which encyclopedia produced each version. Wikipedia averaged 3.9 errors per article. Britannica averaged 2.9. Both had four "serious errors," misinterpretations of important concepts. "Wikipedia comes close to Britannica in terms of the accuracy of its science entries," Jim Giles reported.[101]

Britannica fired back with a 20-page rebuttal titled "Fatally Flawed," calling Nature's method incompetent.[102] Nature stood by its findings. The academic consensus shifted anyway. Wikipedia was credible.[137,138] The open model worked. Defenders had a further point. Britannica made errors too, and the difference was that Wikipedia could fix them.

Then the restrictions came. On December 5, 2005, Jimmy Wales announced that anonymous users could no longer create new articles on the English Wikipedia. "As an experiment, we will be turning off new pages creation for anonymous users," he wrote. He stressed his continued belief in anonymous editing: "I am a firm believer in the validity of allowing anons to edit."[139] But the line had moved. "Anyone can edit" now came with an asterisk.

The Biographies of Living Persons policy tightened. Contentious unsourced material about living people would be removed at once, without waiting for discussion. A new function called oversight let a small group of trusted users hide content from the revision history entirely, making libel invisible even to other editors.[47]

The machinery worked because someone chose to build it. The vandal, Brian Chase, was identified not by Wikipedia or law enforcement but by Daniel Brandt, an outside critic who traced the IP address to a Nashville delivery company.[140] Chase was never prosecuted. No clear criminal law applied. By then the libel had spread to Answers.com and Reference.com, propagating through automated content sharing faster than any correction could follow.[4]

The deeper lesson was about attention, not anonymity. The contradiction between Seigenthaler and Nature was more apparent than real. Both were right about different things.

Wikipedia's science articles drew many expert contributors and constant monitoring. Vandalism on "George W. Bush" was fixed in seconds, because the article had many watchers. Seigenthaler's biography, a low-traffic entry about a regionally known figure, had none. The hoax survived

four months because no one was looking. The wisdom of the crowd varied with the size of the crowd watching.[135,4]

The quality crises were being addressed. The enforcement gap was not.

Who Enforces?

The 2009 migration to CC-BY-SA solved the friction problem.[15] It did not address the enforcement problem at all.[136]

Consider who might enforce Wikipedia's license. The Foundation cannot sue on the license that matters most. Copyright in each article stays with the editor who wrote it, and the Foundation holds no assignment, only the same public CC-BY-SA license everyone else receives.[16,141] Under United States law a bare non-exclusive licensee has no standing to sue for infringement, so the Foundation cannot bring a copyright claim over reused article text. The editors can. No one else can. Meta-Wiki says so plainly: "The Wikimedia Foundation owns almost none of the content on Wikimedia sites—it is owned by the individual creators."[136]

The Foundation was not entirely powerless, though, and the gap that leaves is telling. What it *can* enforce is narrower than the commons it hosts: its trademarks, and its own contracts. Anyone who signs for the Wikimedia Enterprise API accepts a binding agreement, with attribution duties, a bar on building a competing service, and injunctive relief for breach, that the Foundation could invoke directly.[142] But that contract governs the paid service, not the free content moving through it, and it binds only the companies that sign.[143] A firm that never signs and simply scrapes the public pages, as the heaviest users did, slips past it entirely.

Even so, over the paying customers it did hold under contract, the Foundation had a lever it never pulled. And the license itself was stronger than a polite request. Attribution and share-alike are conditions of the grant, not wishes attached to it, and breaching a license condition is

copyright infringement, not merely a broken promise.[144] The enforcement problem was never purely structural. It was also a choice.

Creative Commons will not enforce either. "Creative Commons is not a law firm and cannot represent you or give you legal advice." CC writes licenses. It does not sue on behalf of the people who use them.[145]

That leaves individual contributors, for whom enforcement is impractical. A contributor who wrote one paragraph in a 50-paragraph article would first have to identify the violation and verify that their specific content was used. Then they would have to weigh a likely fair use defense—in the United States, an affirmative defense the copier bears the burden of proving, and a doctrine many other countries do not recognize at all—decide whether to sue, bear the costs, and accept that damages might be minimal. For most contributors, that calculus never favors enforcement.[145]

The system assumed good faith. When reusers comply, with proper attribution and share-alike for derivatives, everything works.[16] When they don't, no institutional enforcer exists.[136]

This was not inevitable. The Linux community built enforcement organizations, the Software Freedom Conservancy and the Software Freedom Law Center, that kept enforcing the GNU General Public License (GPL) even as the Linux Foundation took corporate money.[146] Open source kept ideological diversity around licensing. Copyleft purists stayed in the community and in its institutions. The Wikimedia Foundation's professional culture came from tech and nonprofit networks that saw litigation as unseemly. It had no equivalent enforcement constituency.

The system worked tolerably when violations were scattered. Underneath sat a structural problem in commons governance. There was no commons government. The Foundation hosted the platform but disclaimed ownership. Creative Commons wrote the licenses but would not enforce them. Individual contributors could enforce but had no means.[136,145] The commons had rules and no enforcer, a design flaw that would prove catastrophic once extraction became industrial.

Case Law

Courts have addressed CC license enforcement, establishing both that the licenses work and that they have limits.[145]

In Curry v. Audax (Netherlands, 2006), Adam Curry, the former MTV video jockey, posted photos on Flickr under CC-BY-NC-SA, a variant that also barred commercial reuse. A Dutch tabloid, Weekend, published them without attribution. The court ruled that "Audax has not observed the conditions stated in the License" and set a fine of 1,000 euros per photo for future violations. It was the first court enforcement of a Creative Commons license, and it proved the licenses were legally binding.[147,148]

In Drauglis v. Kappa Map Group (US, 2015), a photographer licensed images under CC-BY-SA. Kappa Map used one on an atlas cover. The photographer argued the atlas was a derivative work that required share-alike licensing. The court disagreed: the atlas was a "collective work," not a "derivative work,"[149] and share-alike applies only to derivatives. The ruling narrowed copyleft's reach. Including CC content in a larger work does not automatically trigger share-alike.[16]

In Philpot v. Media Research Center (US, 2018), a photographer licensed images under CC-BY, and the Media Research Center used them without attribution. The violation was clear. The court found fair use anyway. The use was "transformative," concert photos repurposed for political commentary, and fair use defeated the license claim. A violation could still be non-infringing if the use transformed the work.[150]

Trolling Paradox

Some photographers found a business model in the enforcement gap. They posted images under CC licenses, used automated tools to find violations, sent demand letters threatening statutory damages (penalties copyright law imposes regardless of real damage), and settled for payments that exceeded any actual harm.[151]

Marco Verch, a German photographer, sent more than a thousand automated demand letters and filed dozens of lawsuits over minor attribution failures.[152,153] One of his attorneys, Richard Liebowitz, was labeled a "copyright troll" by American judges.[153,154] The strategy made money from people who thought they were using "free" images. Verch was a symptom, not an aberration. When the institutions that issued the licenses declined to enforce them, the vacuum filled with the least idealistic enforcers available.

Creative Commons responded bluntly in a 2022 statement on copyleft trolls. The "license-enforcement-as-business model," it wrote, was "a perversion of the founding ideals of Creative Commons."[145] The statement created a paradox. Aggressive enforcement was discouraged as harmful to trust in CC-licensed content. Passive enforcement let widespread noncompliance persist.

CC-BY-SA 4.0 includes an automatic cure provision. Violations fixed within 30 days of discovery reinstate the license. This favors correction over punishment. It also means violators face little risk if they simply add attribution once caught.[16]

What AI Changes

Every enforcement mechanism assumes human reuse. A photo on a website can be found, identified, and compared. An article copied to another site can be compared line by line and traced. A book that includes Wikipedia excerpts can be examined. When a large language model (LLM) trains on Wikipedia, none of these assumptions hold.[69]

Is the model a "derivative work"?

CC-BY-SA requires that "Adapted Material," meaning derivatives, be released under the same license.[16] Is a trained model a derivative of its training data? The industry says no. Training turns text into statistical weights, and weights, it argues, are not a modified copy of anything.

Look at what training optimizes. A language model minimizes next-token loss, a score that rewards predicting the exact next word of its training text. Minimizing that loss means maximizing reproduction fidelity. The better the model, the more precisely it can reconstruct what it read. And reconstruct it does. In 2021, Carlini and colleagues showed that models emit training text word for word when prompted the right way.[9] Nasr's team later pulled memorized data out of ChatGPT in bulk.[10]

Verbatim output carries a statutory consequence. Under the Copyright Act, a copy is any material object in which a work is fixed. What makes it a copy is that the work "can be perceived, reproduced, or otherwise communicated ... with the aid of a machine or device."[155]

If a prompt makes a model recite a Wikipedia passage word for word, that passage is fixed in the weights and reproducible with the aid of a machine. The weights hold a copy. A new work built by copying that expression, and carrying it forward, is a derivative under the same definition.[155] Share-alike attaches to derivatives, and the exclusive right to reproduce a work belongs to its owner, so both are implicated.[16,156]

This is a serious argument, not settled law, and no court has tested it head-on. The counterargument is real. AI companies call the trained model a statistical artifact, not a copy, and call training transformative rather than reproductive.[157] In Kadrey v. Meta (US, 2023), a federal judge dismissed as "nonsensical" the claim that the models were themselves infringing derivative works.[158]

Transformation describes a purpose. It does not deny that copying occurred. Verbatim reproducibility undercuts the cleaner defense that no copy exists at all. No amount of process complexity changes the statute. A work reproducible from the weights with the aid of a machine is fixed there, however opaque the machine that holds it.[155]

The legal weight of this belongs to later chapters. Chapter 21 works the fair-use factors, and Chapter 14 lays out the memorization evidence. Here the point is narrower. Share-alike was written for recognizable

modifications, and legal scholars Kacper Szkalej and Martin Senftleben concluded in a June 2024 report for Open Future that such licenses are "largely ineffective" against model training.[69] Whether that reflects the law or only the absence of anyone willing to test it is the question the Foundation never forced. When Jillian Bommarito's team asked the Foundation in 2024 how to comply, its staff conceded that many companies were not, in letter or in spirit.[18]

Where is attribution in generated text?

Traditional attribution lives in captions, credits, and metadata. When ChatGPT answers from memory, drawing on what it absorbed in training rather than a live web search, no Wikipedia link appears. No editor names are credited. No CC-BY-SA notice is shown. The user cannot know the source. The license requires attribution "reasonable to the medium." What is reasonable for an AI chatbot? The license does not say.[16]

Is the output share-alike?

If AI output is shaped by Wikipedia training, must it carry a CC-BY-SA license? Even scholars sympathetic to open licensing doubted it. "It has been argued that all outputs generated by an AI are derivatives of all of the inputs, but this argument holds no water," wrote Andres Guadamuz, a copyright scholar at the University of Sussex, on his TechnoLlama blog.[159] When you ask a chatbot to write a poem, he noted, the words are not a derivative of any particular training input.

The Exception Override

CC-BY-SA 4.0, Section 2(a)(2), contains a provision that may render all of this moot: "For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions."[16]

If AI training qualifies as fair use in the US, or text and data mining in the EU, the license conditions do not apply at all.[16] No attribution. No share-alike. The US Copyright Office's May 2025 report concluded that some AI training may qualify as fair use. It warned, though, that commer-

cial training on vast troves of works that compete with the originals would exceed established boundaries.[160] Judge Alsup, in Bartz v. Anthropic (June 2025), called AI training "transformative—spectacularly so."[157]

If that view prevails, CC-BY-SA becomes unenforceable against AI companies. Not because the license is weak, but because the use falls outside copyright's domain entirely. An unenforced license had been tolerable when violations were scattered and harm was diffuse. It was not tolerable when a handful of companies trained on the entire corpus. The legal armor meant to protect the commons becomes irrelevant, the commons open, no walls at all. But that view still had to win in court, and it was far from settled.

License Without Enforcement

Wikipedia fixed one gap and left the other open. Seigenthaler's biography was corrected, Brian Chase apologized, and the biographies policy helped prevent future incidents of similar duration.[4] Wikipedia read that crisis as a quality problem and answered it with quality policies. License compliance got no comparable attention. Between roughly 2005 and 2012, the encyclopedia validated its accuracy, qualified its openness, and built the corpus AI would later extract, every article and every revision accumulated under CC-BY-SA.[15] It also established a habit of non-enforcement that no one ever broke.[136]

The copyleft armor Stallman had forged after Symbolics worked only when someone was willing to wear it, when the enemy was recognizable and the battlefield familiar.[2] When the enemy became statistical weights and the battlefield became fair use doctrine, the armor sat unworn. The Foundation that inherited the framework had never prioritized enforcement, and it would spend the next decade growing without ever building the capacity. By the time industrial extraction arrived, the enforcement capacity still did not exist. The copyleft promise was made. The enforcement never followed.

Part III

The Institution's Turn

CHAPTER NINE

The Foundation

> Servers cost money. Servers eat electricity. Servers need repairs. Bandwidth costs money.
>
> — Wikipedia fundraising discussion, October 2003

On June 20, 2003, Jimmy Wales filed six pages of documents with the State of Florida.[100]

The Articles of Incorporation were bureaucratic boilerplate: a nonprofit corporation, formed for educational and charitable purposes, headquartered in St. Petersburg. One incorporator: Jimmy Wales. The filing fee was minimal. The paperwork was routine.

But these six pages created the legal vessel for what became the world's largest encyclopedia.[47] Every article Wikipedia's contributors wrote, every edit and image and reference, fell under the stewardship of the entity born that day. The Wikimedia Foundation became the institutional home of free knowledge.[100]

The announcement to the community was brief. Wales explained that he was transferring the project's domain names and the copyrights Bomis held to the new nonprofit. The contributors who wrote the articles kept their own copyrights, licensing their work to anyone under the project's free license. The servers, for the moment, stayed with Bomis, with donation to follow once the tax implications were clear. The Foun-

dation was meant to be an "independent charitable entity," separate from Bomis.[100] The community had worried about commercial pressures, and the Foundation was supposed to protect against them.[63]

Told this way, the founding is a story of generosity: a founder building an institution to serve the commons.

The story was true. It was also incomplete.

Bomis Subsidy Era

Wikipedia launched in January 2001 under the corporate umbrella of Bomis, Inc., Jimmy Wales's for-profit company.[47,44]

Bomis provided everything: servers, bandwidth, technical support, Larry Sanger's salary. The project existed because Wales funded it from his company's revenue.[36,4] That revenue had an uncomfortable origin. Bomis ran a web portal that included "Bomis Babes," adult content that, while not pornographic, served what Wales later described as "a market similar to say Maxim magazine... kind of a guy-oriented search engine."[44,161]

The community knew about this background and was broadly unbothered. What mattered was that Wikipedia existed and was growing. By 2003, the project had over 100,000 articles in English and was expanding to dozens of other languages.[47]

But Bomis was struggling financially. The dot-com crash had hit hard. The company could not subsidize Wikipedia indefinitely.[4,36] And the community was increasingly opposed to the obvious alternative: advertising.[63]

Why Incorporate?

A nonprofit foundation solved multiple problems:

Legal liability. Who was responsible for Wikipedia's content? Bomis? Wales personally? The contributors? A foundation with clear legal status would provide protection.

Trademark control. The "Wikipedia" name had value. Someone needed to hold it formally.

Tax-deductible donations. Readers who wanted to support Wikipedia couldn't get tax deductions if they sent money to Bomis. A 501(c)(3), the U.S. category for tax-exempt charities, would enable charitable giving.[67]

Organizational credibility. Wikipedia aspired to be a serious reference work. Operating under an adult-content company complicated that aspiration.

Independence. The community expected separation from commercial interests. A nonprofit would formalize that separation.[63]

Asset Transfer

Wales's decision to create the Foundation was genuine charitable giving.

He transferred the domains (wikipedia.org, wikipedia.com) and the copyrights Bomis held at the founding. The trademarks and server hardware followed.[100] He kept nothing for himself. A less generous founder could have structured things differently, keeping the trademarks, licensing them back, holding control through ownership rather than governance.

Wales did not do this. By the time Wikipedia reached its millionth article across all languages in September 2004, Bomis had invested roughly $500,000 in the project.[47,4] Wales walked away from all of it.

The generosity was real. The structure it left behind was another matter.

Board Structure

Wales incorporated the Foundation as sole incorporator. In January 2004, he appointed Tim Shell and Michael Davis as additional trustees, his Bomis co-founders.[162,100]

The initial board was entirely Bomis. No community representation existed. No mechanism for community input had been created. The people who built Wikipedia's content, a growing army of volunteer editors, had no voice in governing the institution that now held their work.[162]

This was not unusual for nonprofits. Founders typically control the organizations they create. But Wikipedia was not a typical nonprofit. It was a collaborative project where the value came from contributor labor. The governance structure gave those contributors no formal role.

This mattered because the Foundation was not a company managing its own product. It was a steward of assets contributors had built.

First Fundraiser

In October 2003, in the Foundation's first months, Wikipedia's servers went down. The crisis triggered the Foundation's first fundraising campaign.[100,23]

The appeal was blunt: servers cost money, and Wikipedia needed help paying for them.

The goal was $20,000. Within a week, the community had donated over $30,000. By year's end, the Foundation had raised $56,666, enough to purchase nine new servers and begin building distributed infrastructure.[23,100]

The success proved something important: volunteers would financially support what they had built. Wikipedia could sustain itself through small donations from readers and contributors. It didn't need advertising or commercial partnerships.

The modest fundraiser established a model. It also established a dependence on donation revenue, one that deepened as the institution grew far beyond what the 2003 donors imagined.

Community Representation

By 2004, concerns about the all-Bomis board had emerged. The community that wrote Wikipedia wanted a voice in its governance.

The Foundation responded with board elections. From May 30 to June 12, 2004, eligible Wikipedians voted for two trustee positions. Requirements were minimal: a 90-day-old account with editing activity. Angela

Beesley and Florence Nibart-Devouard, both longtime volunteer editors, won.[163,162]

This was progress. The community now had representation. But the structure remained tilted: three appointed trustees, two elected. The appointed members held the majority.[162]

This ratio, appointed majority over elected minority, persisted. Even as the board expanded, the pattern held: those who built the content had voice but not control.

Tax-Exempt Wait

The Foundation operated for nearly two years without formal tax-exempt status. The IRS determination letter arrived in April 2005, retroactive to June 20, 2003.[67]

What this delay revealed: the Foundation's limited initial capacity. This was not a sophisticated institution with professional staff and legal teams. It was a handful of volunteers managing paperwork for a rapidly growing website.[100]

But it revealed something else. The institution was already making decisions about governance, strategy, and Wikipedia's direction before it held any formal charitable status at all.

The Absent Co-Founder

The 2003 Foundation was tiny: about $57,000 in assets, no paid staff, volunteers answering emails.[23] It grew into something more than a thousand times larger. That growth was not corruption. Governing a top-ten global website demanded infrastructure, legal defense, technical work, and community coordination.[164] Volunteers alone could not scale to meet the work.

But the architecture set in 2003 shaped everything after it. The clearest sign came in 2008, when the board formalized a Founder's Seat that no one but Jimmy Wales can hold. Not even Larry Sanger, who designed

Wikipedia's core policies and served as its first editor-in-chief, qualifies.[165,36] The bylaws encoded what incorporation and funding had already decided: governance followed ownership, not contribution.

The seat gave institutional form to a quarrel the two men never resolved. As Chapter 3 recounts, Sanger had claimed co-founder standing since he left in March 2002, pointing to the project's name and its Neutral Point of View policy.[55] Wales tied the title instead to ownership of the assets he had funded and donated.[4] What the Founder's Seat added was an institutional consequence: a chair defined so that the man who wrote Wikipedia's founding rules could never occupy it.

The person who designed the rules was shut out of administering them. It was the early shape of a pattern: an institution created for the community but not by the community. Wikipedia's value came from collective labor, not just Wales's money or Sanger's policies but a quarter million contributors adding knowledge article by article. The institution governing that work was not built by the collective. Its structure reflected legal ownership, and the fight over a title was, at bottom, a fight over who controlled the result.

Servers Cost Money

The October 2003 server appeal worked because donors believed they were keeping the site online.

The institution those donors funded grew until the honesty frayed. In 2022, English Wikipedia's editors took up the question in a formal Request for Comment (RFC), the site's process for testing community consensus. They reached near-unanimous agreement that the fundraising banners had become misleading: urgent appeals for survival masked an organization sitting on years of reserves.[166,167] What began as a plea for server money had become a fight over institutional reserves.

The Foundation was built to act as a trustee, guarding the shared resource its contributors had assembled. Incorporation left one question

open: whether that trustee could stay faithful once it developed its own survival imperatives.

CHAPTER TEN

The Fundraising Machine

> Imagine a World Without Free Knowledge.
>
> — Wikipedia blackout page, January 18, 2012

At 05:00 UTC on January 18, 2012, the English Wikipedia went dark.[168]

In place of the familiar white page with its blue links and articles, visitors found a black screen with white text. No search box. No navigation. Just a stark message and a tool to look up their representatives by zip code:

For over a decade, we have spent millions of hours building the largest encyclopedia in human history. Right now, the U.S. Congress is considering legislation that could fatally damage the free and open Internet. For 24 hours, to raise awareness, we are blacking out Wikipedia.[169]

The bills were the Stop Online Piracy Act (SOPA) and the PROTECT IP Act (PIPA). Both would have let content owners force internet providers to block sites accused of hosting infringing material. Wikipedia hosted millions of images and text excerpts. Accusations against it were inevitable, which made the project a plausible target.[168]

Some 162 million people saw the black page. Eight million looked up their congressional representatives. Congressional websites overloaded. Within days, legislation that had seemed inevitable weeks earlier was shelved indefinitely.[170,169]

A volunteer project had mobilized political power at scale. The world's sixth-most-visited website turned its reach toward a cause and won.[171]

It was the peak of Wikipedia's public power.

Survival to Surplus

The Foundation that raised $57,000 in 2003 had become something else entirely.[23]

In 2007, when Sue Gardner arrived as executive director, the Wikimedia Foundation had roughly ten employees and $2.7 million in annual revenue.[23] Wikipedia ran on servers that crashed regularly. The organization operated from St. Petersburg, Florida, far from the tech industry's centers of power.[47]

Gardner had run CBC.ca, one of Canada's leading news sites.[172] She understood how to grow organizations. Within four years, revenue passed $25 million and the staff grew to about fifty.[23] Within seven, the Foundation's own materials called it the fastest-growing nonprofit in America by revenue.[172]

The model was simple: banner appeals asking for small donations. The execution was sophisticated. A/B testing measured every element. Colors, wording, suggested amounts, placement, all of it was tested. The banners that performed best were deployed. The rest were discarded.[173]

"Wikipedians do it for love, for mission-type reasons," Gardner explained. "They don't want to be paid. They want to be praised."[174]

The fundraising machine had learned to convert that love into revenue.

Wales Effect

In November 2010, the Foundation found its most powerful fundraising tool: Jimmy Wales's face.[175]

"Please Read: A Personal Appeal From Wikipedia Founder Jimmy Wales."[175,173]

The banner showed Wales looking directly at the camera, his expression earnest.[175] The personal appeal was new to encyclopedia fundraising, which usually emphasized institutions rather than individuals. This was a named person, making eye contact, asking for help.

The results were striking. The 2010 campaign was the shortest fundraiser in Wikimedia history, and it raised $16 million.[173] The Wales banner outperformed every alternative by a factor of fifteen.[176]

The banner also became a meme. "Jimmy Wales's creepy stare" generated parodies across the internet.[177,175] The Foundation did not mind. Whatever drove people to talk about Wikipedia also drove them to donate.

Later testing revealed something curious. The personal appeal worked regardless of who appeared. Brandon Harris, a Foundation designer, became another effective face.[178] The magic was in the format, the direct address and the sense of personal connection, not the specific person.

But Wales remained the primary face. The founder had become the fundraising machine's most valuable asset.

The Decision to Go Dark

The Stop Online Piracy Act moved through Congress in late 2011. Critics warned it would break the internet's architecture and enable censorship.[168] The question for Wikipedia's community was sharper. Should an encyclopedia take a political position at all?

From December 2011 to January 2012, the English Wikipedia held a formal request for comment. Roughly 1,800 editors weighed in on a single question. Should Wikipedia black out in protest?[168]

Defenders argued self-defense, not advocacy. SOPA threatened Wikipedia's operations directly. Sites could be blocked on accusation alone, without due process. A project hosting millions of images and excerpts could not avoid such accusations.[168]

Gardner reached for an analogy she liked. "Wikipedia is like the National Parks Service," she would say. "The Internet is a vast space and

it will only continue to grow, but in the vastness you still need space for parks or public libraries."[174] Protecting the public library, in this view, was not politics. It was survival.

A minority warned about precedent.[168] Wikipedia is a reference work, not an advocacy organization. Taking political sides, even popular ones, could compromise the trust that made the project valuable. The Neutral Point of View policy existed precisely to keep Wikipedia out of such disputes.[96] The blackout would work because the cause was popular. What happens, the dissenters asked, when the cause is controversial? Who decides which causes are legitimate?

Gardner acknowledged the tension before the blackout began. "In making this decision, Wikipedians will be criticized for seeming to abandon neutrality to take a political position," she wrote. "That's a real, legitimate issue. We want people to trust Wikipedia, not worry that it is trying to propagandize them."[179]

The community decided anyway. About 90 percent of the editors who commented supported the blackout. Google, Reddit, and thousands of other sites joined with protests of their own.[168] The Electronic Frontier Foundation counted over a million emails to Congress.[170]

When the protest ended, Gardner spoke for the project. "The Wikipedia blackout is over and the public has spoken," she announced. "You said no."[180]

Shadow of the Harvest

The blackout was the high point of Wikipedia's political power. A volunteer community had mobilized hundreds of millions of viewers and helped kill legislation backed by major entertainment corporations.[168]

It was also the turn. Not toward failure as an encyclopedia. Wikipedia would keep growing. The turn was toward a different kind of institution. The steward that had proven it could defend the commons would soon

learn that defense requires choosing enemies. Choosing enemies becomes hard when the enemies become customers.

The harvesting began the same year. In May 2012, Google launched Knowledge Graph.[181] The product displayed Wikipedia content directly in search results, information boxes that answered questions without sending users to Wikipedia. The content was extracted and shown where Google could monetize the attention.[182]

The year Wikipedia proved it could beat a legislative threat, a different threat was born. This one would not come through Congress. It would come through APIs and training data and AI systems that used Wikipedia without sending traffic back.[21]

The Power Left Unused

A decade after the blackout, critics asked a different question. If Wikipedia could go dark to defeat SOPA, why could it not go dark to make AI companies comply with its license?

By 2024, AI companies were training on Wikipedia at scale.[18] Chatbots leaned on it more heavily than on any other single source, with no payment flowing back and little attribution reaching users.[12,183] The Foundation's own human traffic was falling while automated reuse grew.[21] SOPA had threatened the project through law. The harvesting came through technology.

The Foundation answered AI ingestion not with blackouts but with Enterprise contracts.[29] It sold access to the same companies whose compliance with Wikipedia's license was in question.[183]

The capacity for collective action still existed. It was exercised exactly once. The community that blacked out the internet never blacked out again. Not when Knowledge Panels absorbed its content without a click-through.[182] Not when voice assistants read its answers without attribution.[184] Not when AI companies trained on the entire corpus and built products worth billions.[18]

The difference was not capability. It was will. And will, inside an institution, follows incentives. By 2024 the Foundation drew revenue from AI companies through Enterprise,[20] and it held more than $400 million in reserves and endowment.[23,185] Institutions rarely threaten the companies that sign their checks.

That collective power sits unused. It did not disappear. The institution that could wield it had grown dependent on the revenue the extraction produces. What had changed were not its powers but its incentives.

CHAPTER ELEVEN

The Apparatus

> Perhaps it's time to stop calling self-selected surveys of a tiny subset of our user base "community consensus."[186]
>
> — Brion Vibber, WMF Lead Software Architect, 2014

On January 31, 2008, the Wikimedia Foundation closed its Florida headquarters.[187]

A handful of employees packed their belongings and headed west. They were leaving St. Petersburg behind, a quiet nonprofit home far from the tech industry. San Francisco was the destination: ambitious, expensive, and connected to Silicon Valley's employers, funders, and professional networks.[188]

The stated reasons were practical: a deeper talent pool, proximity to partners, cheaper international travel.[188] The unstated aspiration was evident. Wikipedia was joining the tech industry.

San Francisco was not just expensive and ambitious. It was also among the most politically homogeneous major cities in America, a Democratic bastion where professional networks sorted by ideology as surely as by skill. Hiring from these networks would shape who staffed the institution.

What gets left behind in a move is not just geography. It is identity. The scrappy volunteer project that ran on donated servers and volunteer labor

was becoming something else: an institution with office space, employee benefits, and strategic planning documents.[4]

The move to San Francisco was the visible symbol of a transformation that would take fifteen years to complete. When it finished, the volunteer project had become an institution that negotiated enterprise contracts with Amazon, Google, Meta, and Microsoft.[19]

The apparatus those employees were about to build would answer to interests its founders never imagined.

Building the Apparatus

The move to San Francisco was Executive Director Sue Gardner's instrument. The fundraising machine had already given the Foundation money. Now Gardner had to build the institution that money implied: staff, infrastructure, and process.

By the time she stepped down in mid-2014, the Foundation employed about 200 people on a budget above $50 million.[189,23] Reliable infrastructure had replaced the servers that once crashed weekly. Forbes ranked her among the 100 most powerful women in the world.[190] These were real accomplishments.

But professionalization was more than fundraising and uptime. It was the construction of an apparatus, and the apparatus would soon develop interests of its own.

Revenue Trajectory

The fundraising success was remarkable, as shown in Table 11.1.

Revenue grew 68-fold.[23] Staff grew from roughly ten to seven hundred.[191] The organization became more labor-intensive even as its primary product, the volunteer-created content itself, required zero Foundation employees to produce.

Year	Revenue	Staff	Revenue per Staff
2007	$2.7M	10	$270K
2011	$25M	50	$500K
2014	$52M	200	$260K
2020	$129M	450	$287K
2024	$185M	700	$264K

Table 11.1. Wikimedia Foundation revenue trajectory, 2007–2024

These employees came from somewhere, recruited through particular professional networks. Those hiring pipelines shaped the Foundation's defaults about whom to challenge and whom to treat as partners.

The growth was not irrational. Running one of the world's most-visited sites takes real staff: servers, lawyers, coordinators, planners.[171] Someone has to do this work. The question is how much, and who decides.

Staff Over Consensus

Between 2013 and 2019, four governance crises exposed the same fault line. Professional staff, pursuing strategic goals, repeatedly overrode the volunteer consensus the Foundation was built to serve. Each crisis ran the same course.

Staff pursued an objective. The community objected through its established processes. Staff dismissed the objection as unrepresentative. The community resisted. Staff eventually backed down, but only after trust had eroded. The four episodes used different mechanisms. Each proved the same drift.

The first crisis began with a genuine problem. Wikipedia's editor population was declining. Between 2006 and 2011, the English Wikipedia lost roughly 15,000 active editors. Studies pointed to wiki markup, the technical syntax editors used to format articles, as a barrier to newcomers.[192] The Foundation's answer was VisualEditor, a tool that let editors

format text by clicking buttons instead of learning markup. A $3.6 million grant from the Stanton Foundation funded the work.[193] On July 1, 2013, the Foundation deployed VisualEditor as the default editing interface, still trailing hundreds of open bug reports.[194]

The community objected. A request for comment showed 212 to 20 consensus to make the tool opt-in rather than default.[195] Staff refused. You could not, they argued, simply add up the numbers and call that consensus. That attitude would recur. On September 23, 2013, an administrator named Kww deployed code to override the Foundation's decision. Within days, the Foundation backed down, and VisualEditor became opt-in.[196]

The next crisis hard-coded staff power into the software itself. The German Wikipedia had voted to disable Media Viewer, a new image-display feature, by default. The Foundation responded on August 10, 2014, with Superprotect, a technical permission that let staff lock pages against community edits.[197] Erik Moeller, the Deputy Director, was blocked on German Wikipedia for edit warring.[186]

This was not a software dispute but a governance one. The Foundation had built a mechanism to override consensus by force. A protest letter drew 824 signatures.[198] A vote went 664 to 103 to remove the permission.[199] Superprotect stayed in place for fifteen months. The Foundation removed it on November 5, 2015,[200] after Executive Director Lila Tretikov acknowledged that it had "set up a precedent of mistrust, and this is something it was really important for us to remove."[201]

The third crisis turned on secrecy. In January 2016, details emerged of a Foundation project called Knowledge Engine, billed as "the internet's first transparent search engine" and funded by a $250,000 Knight Foundation grant.[202] The community had not been consulted. The Board had not been fully informed.[203] Trustee James Heilman was removed from the Board.[204] He maintained the removal was punishment for pushing transparency about the project,[205] a connection Wales angrily disputed.[206]

Staff departed in waves. Luis Villa, Anna Koval, and Siko Bouterse left within days of one another, Bouterse resigning with a pointed public statement. Tretikov resigned on February 25, 2016.[202] An organization built on openness had developed its central strategy in secret.

The fourth crisis struck at the community's self-governance. On June 10, 2019, the Foundation's Trust & Safety team, the staff unit that handles harassment and conduct complaints, banned the administrator Fram for one year. It consulted neither the community nor the Arbitration Committee, the body the community had created to handle exactly such disputes.[207] Jimmy Wales called it "a question about our constitutional order."[208]

Twenty-one administrators resigned in protest. The discussion ran to 470,000 words.[208] "I am not willing to serve Jan and the T&S team... under undisclosed new rules and under threat of unappealable sanctions," one resigning administrator wrote.[207] By late July 2019, the active administrator count had dropped below 500 for the first time.[209] In September, ArbCom vacated the ban.[210]

Four crises, three executive tenures, one pattern. VisualEditor unfolded under Sue Gardner, Superprotect and the Knowledge Engine under Lila Tretikov, the Fram ban under Katherine Maher.[191] None of the four was a personal failing. Each director met the same structural bind: the staff held the resources and the legal authority, the community held only its voice. The tension persisted no matter who led. This was mission drift in miniature: an organization built to serve editors had grown priorities of its own, and when those priorities collided with the community's, the staff held the power to win.

Diverging Trajectories

Foundation and community numbers moved in opposite directions. Table 11.2 sets the two side by side.

Metric	2007	2024
WMF Staff	10	700
Active editors (English)	51,000	30,000
Active administrators	1,000	450

Table 11.2. Foundation and community: diverging trajectories

Staff multiplied seventyfold.[191] Active editors declined 41 percent.[192] Active administrators declined 55 percent.[209] The Foundation grew while the community shrank. The institution that volunteers built now pursues agendas volunteers never authorized.

Necessity and Drift

The Foundation's defenders read the record of growth and conflict as necessity. Wikipedia cannot run on volunteers alone. Professional staff provide infrastructure at global scale, legal defense, quality work across languages, and long-term planning, including the endowment that reached $100 million five years ahead of schedule.[211] By this account, Wikipedia still operates because someone built the institution to keep it operating. Without professionalization, it might not exist at all.

The community read the same record as betrayal. The Foundation had been created to serve editors. Instead it came to treat community consensus as optional input: software shipped without consultation, strategies the community never approved, "reader interests" invoked to override editor decisions. "Wikipedia has Cancer," the community member Guy Macon titled a widely read 2017 essay. The cancer, he wrote, was "the explosive growth in spending, a growth rate that far exceeds the growth in content or the growth in users."[212] The mandate had been to keep Wikipedia online. The organization had become something with its own agenda.

The two readings are not rivals. They describe one structure from opposite ends, and that is the point. Professional organizations develop professional priorities. Volunteer communities hold different values. The conflict is not that either side is wrong, but that the two are incompatible. Staff need to plan, hire colleagues, and manage budgets. Editors want to write an encyclopedia without bureaucratic interference.

Whether Foundation employees counted as members of the community became its own debate after VisualEditor. But employees follow professional incentives and volunteers follow mission incentives. When the two diverge, the side holding legal control, money, and staff tends to prevail over the side holding only content and labor.

The Institution Complete

By 2024, the apparatus was complete. The expansion had become the Foundation's normal operating model: an endowment that had hit its targets early, operations running remote-first across the globe.[211]

This was a genuine achievement. The shoestring operation that crashed regularly in 2007 had become a durable institution. Wikipedia would not vanish because a server failed or a key volunteer burned out.

It was also a transformation. The volunteer project had acquired a professional apparatus, and the apparatus had acquired an agenda that pulled away from the community creating the content. A 700-person institution carries an enormous payroll, roughly $80 million in salaries and wages on the Foundation's most recent Form 990, the annual U.S. nonprofit tax filing, before benefits and officer compensation.[213] When donations plateau, new revenue becomes necessary.

The logic ran in one direction. The institution needed to survive. Survival required growth. Growth required revenue. Revenue required Enterprise, the commercial arm that would sell API access.[38] Enterprise required serving AI companies.[29] Serving AI companies required not enforcing licenses against paying customers.

An organization meant to protect a shared resource from commercial exploitation had grown dependent on commercial relationships to survive. The founder control embedded in 2003 left no democratic check on the choice.[165] The editors who created the content had no authority to override it. This was not failure but logical consequence. The arm that would sell access to Google and Amazon did not appear from nowhere.[19] It was built deliberately, funded by small donors who believed they were keeping Wikipedia online.

The move to San Francisco had been an aspiration. The aspiration became an organization. And the organization now faced the choice its own survival had created: defend the copyleft licenses that governed volunteer-created content, or take money from companies whose compliance with them was doubtful.

CHAPTER TWELVE

The Warnings

> The Foundation refrained from taking more decisive, direct action against the concerning group of editors to prioritise safeguarding the principle of community self-governance.
>
> — Wikimedia Foundation, Croatian Wikipedia Disinformation Assessment, 2021

In September 2013, a sitting cabinet minister told his country's students to stop trusting their own encyclopedia.

Željko Jovanović was Croatia's Minister of Science, Education and Sports. Asked what he would say to the pupils and university students who relied on the Croatian-language Wikipedia, he did not soften it. "It is our duty to warn them and all other Wikipedia users," he said, "that a large part of the content on the Croatian version of Wikipedia is not only scientifically dubious but evidently falsified." He urged them toward "more reliable sources of information, including versions of Wikipedia in English and other world languages."[214] A newspaper headline compressed the message to something blunter: kids, don't touch the Croatian Wikipedia, the content is forged.[215]

A government had just warned its own schoolchildren away from a Wikipedia. It was the loudest public alarm any government had yet raised

about the project. The administrators the minister was describing kept their tools for seven more years.

The years between 2012 and 2025 handed the Wikimedia Foundation three full rehearsals for the test that was coming. A community organ was captured by ideologues, and the Foundation watched. Commercial operators abused the encyclopedia for profit, and the Foundation's lawyers wrote a letter. Donor money moved through an outside vehicle, and the disclosure arrived as a footnote. Three different organs, one institution, one reflex.

The Captured Encyclopedia

Croatian Wikipedia was born in a split. In February 2003, the original Serbo-Croatian project divided along national lines, spawning separate Croatian, Serbian, and Bosnian editions.[216] The Foundation's own later assessment identified the design flaw: carving one pluricentric language into small national projects "deprived the newly-created communities of editorial diversity" and let editors sort themselves by politics.[217] A large Wikipedia self-corrects because its editors disagree. A small one built around a national identity can be captured by whoever wants it most.

Someone wanted Croatian Wikipedia very much. In 2009, after a conflict that drove out roughly half the project's active contributors, the bureaucrat SpeedyGonsales and administrator Roberta F. installed two allies, Kubura and Zeljko, as administrators.[218] Kubura had been made an administrator that September over objections citing his nationalism, and in 2017 the diminished community would promote him to bureaucrat by a vote of 21 to 0.[219] By 2011, on the Foundation's own reckoning, the group held "undue de-facto control over the project," and it kept that control until 2020.[220]

What they did with it was rewrite the darkest chapter of Croatian history. The Foundation's assessment found the content distorted "in a

way that matched the narratives of political organisations and groups that can broadly be defined as the Croatian radical right."[220]

Jasenovac, the concentration camp complex where the fascist Ustaša regime murdered Serbs, Jews, and Roma by the tens of thousands, was reclassified as a "collection camp." The article gave a revisionist activist who put the death toll near 1,500 equal footing with the memorial site's documented count of more than 83,000.[221,217] Even Auschwitz was retitled a "collection camp."[221] The article on Ante Pavelić, the regime's dictator, described a "politician, lawyer, leader" where the English edition said "fascist" and "military dictator."[221]

The corruption ran deeper than framing. Investigators at the Croatian outlet Novosti documented fabricated quotations that stood for years: an invented Tito speech contradicted by the digitized archive of the real one, and a Moša Pijade speech sourced to an archival file that does not exist. A contributor flagged the fabrication directly to SpeedyGonsales in 2018. It stayed.[218]

The article on same-sex marriage was rewritten to call it "an oxymoron."[222] Administrators scrubbed sourced, court-documented convictions from the biographies of friendly public figures, at the subjects' own request, and blocked the editor who tried to restore them.[218] By 2018, two of the administrators had personally created roughly 40,000 of the project's 190,000 articles.[218] The historian Hrvoje Klasic told an interviewer he would never send his students there.[221]

Holding power took more than editing. Dissenters were blocked, including an editor removed for taking unsourced propaganda out of the Ustaša article.[219] More than a dozen appeals to Meta-Wiki, the movement's coordination site, died the same procedural death. The Foundation's assessment described the pattern: discussions that gained momentum would "branch out from the original topic," turn personal, and drown in irrelevant comments.[217]

In October 2013, at the height of the national scandal, the local community voted on stripping the three leading administrators of their tools. The administrators survived with about 53 to 55 percent support.[216] Seven years later, a checkuser investigation established that some of those supporting votes had been cast by fake accounts. Of 37 accounts investigated, 19 were likely sockpuppets of Kubura, and the network had tipped on-wiki votes and even a 2020 steward election.[223]

A Decade of Alarms

No one can say the Foundation was not told.

The Croatian press first raised the alarm in February 2012.[216] In September 2013 the story went national: a front-page investigation in *Jutarnji list* under the headline "Right-wingers took over the editing of Croatian Wikipedia," the minister's warning three days later, a protest Facebook page that drew thousands.[224,215] Jimmy Wales told a Croatian daily that year that the Foundation was aware of the problems and looking into them.[217] A formal request for comment opened on Meta-Wiki that same month, ran for roughly six months, and closed without result. The Foundation's own report later supplied the epitaph: "There was no Foundation process in place backstopping the struggling community effort."[217,225]

The alarms kept coming. In 2018, Balkan Insight published "How Croatian Wikipedia Made a Concentration Camp Disappear," and Novosti published its diff-level investigation under the title "Endehapedia," after the Ustaša state's initials.[221,218] A prominent Croatian writer concluded the project could be called, "without being unjust or exaggerating," an Ustasha Wikipedia.[226]

In August 2019, English Wikipedia's own community newspaper, the Signpost, laid the situation before the entire movement and named the missing actor: "Since there is no opposition left, change has become impossible without outside intervention. It isn't coming from the WMF,

though. They know about the issue, but reportedly have no comment."[226] Through all of it, by the Foundation's own finding, the content grew steadily worse, "with more revisionist claims and disinformation inserted in an increasing number of articles each year."[217]

What finally ended the capture was not a warning. It was a paper trail. In late 2020 a volunteer grew suspicious of an account cluster and filed the checkuser request that exposed the sockpuppet network.[223] Kubura was globally banned that November. The remaining administrators were stripped of their tools in March 2021, and in June the Foundation published the external assessment it had commissioned once the ban was already in motion.[219,220] Fourteen years had passed since, by the report's own account, the global community and the Croatian public first became aware of the problem. Eight had passed since a cabinet minister said it on the record.

The report did not describe an oversight. It described a policy: the Foundation had held back "to prioritise safeguarding the principle of community self-governance."[217] Its Trust and Safety chief, Jan Eissfeldt, was candid about the blind spot: the Foundation "did not adequately understand" the risks of splitting a pluricentric language into small national projects.[227] Researchers who later compared the Croatian and Serbian editions, in a study titled "Governance Capture in a Self-Governing Community," found the difference was not exposure to nationalism but structure. The captured project had prized informal trust networks over codified rules and kept newcomers out of governance. Its offices went to the clique that organized first.[228]

Set the two reflexes side by side. When German Wikipedia voted to disable a software feature in 2014, the Foundation built a new technical permission to override its own editors within days.[197] When Croatian Wikipedia spent a decade converting the Holocaust into a "collection camp," the Foundation adhered to its policy of not intervening and monitored the situation.[217] Intervention was always possible. It was deployed

when the threat was to the institution's authority, and withheld, as principle, when the threat was to the commons itself.

The Reflex

The second rehearsal tested a different organ: the legal one.

In 2013, an Austin firm called Wiki-PR was selling Wikipedia itself. Its staff edited the encyclopedia for paying clients through a network of fake accounts. The community's investigation eventually identified 323 confirmed sockpuppets and 84 suspected ones.[229] Executive Director Sue Gardner called paid advocacy editing a "black hat" practice that "violates the core principles that have made Wikipedia so valuable," and closed her statement with a promise of force: the Foundation was "currently assessing all the options at our disposal."[230]

All the options turned out to be a letter. The community banned Wiki-PR that October. In November, the Foundation's outside counsel at Cooley LLP sent a cease-and-desist demanding the firm disclose its accounts and stop editing without disclosure.[231] No lawsuit followed. The Foundation instead amended its Terms of Use in June 2014, after a consultation that drew 320,000 words of commentary, to require disclosure of paid editing.[232] The rule was written. Whether anyone would enforce it against a determined violator remained untested.

The test arrived within fifteen months, and it was worse. In August 2015, volunteers announced they had blocked 381 sockpuppet accounts running an extortion ring inside the encyclopedia.[233] The operation, named Orangemoody after its oldest account, had industrialized a con. Its operators watched the queue where newcomers' draft articles were declined, contacted the frustrated subjects while posing as trusted editors, and charged hundreds of pounds to move drafts into the encyclopedia. Then, months later, came the second call: pay $30 a month, or the article might get deleted.[234] Some victims paid and lost their articles anyway.[235]

This was a protection racket built on the Foundation's own trademark and the community's good name.

The Foundation's response was a blog post. "No one should ever have to pay to create or maintain a Wikipedia article," it said, standing "with the Wikipedia community."[233] The accounts were banned and the articles deleted. No lawsuit was filed. No prosecution followed. The operators were never publicly identified. Unmasking them would have taken the subpoena power of a lawsuit no one filed.[233,234]

Twice in two years, commercial operators had abused the encyclopedia for profit, the second time through conduct with a criminal name. Twice the Foundation reached for the mildest available instrument and stopped. A ban, a letter, a terms-of-service amendment, a statement of values. The legal capacity existed, as a later chapter will show in detail. What the record established was the reflex: when defending the commons required going to court against a commercial abuser, the Foundation did not go.

The Footnote

The third rehearsal involved no captors and no con men. It was about how the institution moved money, and how it told the people who gave it.

In its fiscal year ending June 2020, the Foundation made an "unconditional grant" of $8.7 million to Tides Advocacy, an outside progressive advocacy nonprofit.[236] About $4.2 million of it was routine funding for Wikimedia's own affiliate organizations, routed through the outside vehicle for administrative convenience. The other $4.5 million launched something new: a Knowledge Equity Fund that would make grants to external organizations working on racial equity in knowledge, most with no operational connection to Wikimedia projects.[237,238] The Foundation's new general counsel had come to Wikimedia from running Tides Advocacy.[236]

No press release announced the transfer. No dedicated Board resolution marked it. It surfaced in a question-and-answer page attached to the

annual audit.[236] A volunteer named Yair Rand spotted it there in December 2020 and posted to the movement's mailing list: a Google search for the fund's name returned nothing that predated the audit's release. "It is clear that the WMF kept this significant move completely secret for over five months," he wrote. "I am appalled."[239] A Foundation staffer pushed back in the same thread, calling it "disingenuous" to describe as secret something acknowledged in the audit's own FAQ.[239] Both statements were true. That was the problem. Disclosure had been engineered to be technically real and practically invisible.

For two years the fund stayed a community grievance. Then the donors found out. In October 2022 a Twitter thread went viral, pairing Wikipedia's fundraising banners with the fund's grants. "If you're like me, you may have donated as a result," it read. "But I've now learnt the money isn't going where I thought."[240] A donor who cancelled a recurring gift received a reply calling the thread "misguided" for failing to reflect "an accurate understanding of what it takes to sustain a top global website."[240]

The Signpost checked the accompanying claims against the Foundation's own audited financials and found the numbers pointing the other way: in the year in question, actual community grants ran to a few million dollars against a $50 million surplus.[240] The head of the nonprofit that runs Wikipedia's university outreach put the community's verdict plainly: "I would encourage you to stop."[240]

The Foundation did not stop, but it adjusted, one concession per year of pressure. In 2023 it announced the fund would move back from Tides Advocacy into the Foundation itself.[241] That October, English Wikipedia editors weighed a non-binding resolution asking the Foundation to return the fund's remaining money and to seek community approval before making grants a reasonable person would consider unrelated to Wikimedia. The vote ran three to two in favor.[242] Later rounds added community nomination and, finally, grants explicitly paired with Wikimedia affiliates. The fund spent down its $4.5 million and closed in July 2025.[243]

Nothing about the episode was corruption. The grantees were real organizations doing documented work, and the Foundation could argue, and did, that the fund served the movement's official strategy.[237] What the episode rehearsed was a pattern, and the pattern is the point. Money was raised under one framing, moved under another, and disclosed in the fine print. When the gap surfaced, the first response was defense, the second was process, and the reform arrived years later, under pressure, in installments. The people who wrote the checks, and the volunteers whose work inspired them, learned the details last.

Three Rehearsals

Alone, none of the three stories decided anything. Together they mapped an institution.

Each response had a respectable rationale. Community self-governance is a real principle. Litigation is expensive and uncertain. Grantmaking discretion is normal for a foundation of its size. That is what made the pattern durable. At every decision point, the defensible option and the passive option were the same option.

By the middle of the 2020s, anyone who had watched closely knew three things about the institution that held the keys to the commons. It would not intervene against capture until forced. It would not litigate against commercial abuse at all. And it would tell its donors and volunteers what it was doing with their money after the fact, if they read the footnotes. The largest extraction in the encyclopedia's history was about to begin, and the institution it would meet had already shown its reflexes.

CHAPTER THIRTEEN

Enterprise

> Our content is there to be used. It's freely-licensed and it's freely-licensed for a reason. At the same time, it's like the environment. It's there to be used, but it's not there to be exploited.
>
> — Lisa Gruwell, WMF Chief Revenue Officer, 2018

In January 2020, the Wikimedia Foundation created a company.[244]

It was called Wikimedia, LLC, and it was incorporated quietly in Delaware.[244] The Foundation chose Delaware for its settled corporate law. The structure walled the nonprofit off from commercial liability while letting a new arm earn money. What the Foundation had built was a for-profit subsidiary.[38]

No announcement marked the moment. The operating agreement was not executed until July 2020. The public announcement waited until March 2021.[245] For more than a year, the commercial arm existed mostly on paper, known to almost no one.

The fundraising banners told a different story. They still asked readers for small donations to keep the encyclopedia free.[246] The pitch to the public was charity. The structure being built in Delaware was commerce.

A project founded on "free as in freedom" was assembling a business to sell premium access to its own content.[107,38] Among the eventual buy-

ers would be companies the Foundation itself suspected of breaking the licenses that made Wikipedia free.[18] The question was how a copyleft encyclopedia became their supplier.

The Steward's Dilemma

The Foundation faced a real problem. For years, tech giants had extracted enormous value from Wikipedia while contributing almost nothing.[247] This broke the basic principle of the commons: those who benefit should help maintain what they benefit from. But solving it meant the steward becoming a commercial actor. That was the very thing it had been created to prevent.

Extraction Without Contribution

The tech giants' extraction was visible across search and voice products.

In 2012, Google launched Knowledge Graph, the engine behind the information panels that appear beside search results. It displayed Wikipedia content directly in Google's results.[182] Users got their answers. Google got advertising revenue. Wikipedia got nothing.[247]

Amazon's Alexa launched in 2014, answering millions of daily queries by reading Wikipedia content aloud.[248] Ask "Who was Albert Einstein?" and Alexa recited Wikipedia's article. No attribution. No compensation. No link for users to follow.[247]

Apple's Siri worked the same way. Voice assistants across the industry treated Wikipedia as a free backend, pulling content to power products worth billions.[247]

By 2018, TechCrunch investigated corporate contributions to Wikipedia. Amazon had contributed $0. Apple, Microsoft, and Facebook had each given about $50,000, mostly through employee matching programs. Google had contributed roughly $1 million, approximately 0.001% of its revenue. The trillion-dollar companies built products on volunteer labor and gave back almost nothing.[247]

Gruwell's environmental metaphor was apt.[247] Like the environment, Wikipedia was being exploited without consequence.

Disintermediation

The free-riding was not only financial. It was structural.

When users receive Wikipedia content through Google Knowledge Panels or voice assistants, they never visit Wikipedia. They never see the "edit" button. They never see the fundraising banners. They never meet the community.[38]

This severed the volunteer feedback loop. Wikipedia's model depended on readers becoming editors.[249] If readers never reached Wikipedia, where would the next generation of editors come from?

Traffic data confirmed the worry. Knowledge Panels, AI Overviews (Google's AI-written answer summaries), and voice assistants all reduced click-through to source sites.[250] Wikipedia was training the systems that competed with it for attention.[21]

The First Compromise

The Foundation had crossed the partnership line once before, and the episode was instructive.

In October 2011 it announced Wikipedia Zero, a program that paid nothing and charged nothing: mobile carriers in developing countries agreed to exempt Wikipedia from data charges, making the encyclopedia free to read on a prepaid phone.[251] The first partner network went live in Uganda in April 2012, and over six years the program spread to 97 carriers in 72 countries, within reach of an estimated 800 million people.[252,253]

The catch was the principle it bent. Zero-rating privileges one website over the rest of the internet, which is precisely what network neutrality forbids, and the Foundation had long counted itself on neutrality's side. Digital-rights groups, including the Electronic Frontier Foundation, said so publicly.[254,255] The Foundation's answer, delivered by Gayle Karen

Young, its chief culture and talent officer, to the Washington Post in 2014, deserves to be quoted for its structure rather than its content: "Partnering with telecom companies in the near term, it blurs the net neutrality line in those areas. It fulfills our overall mission, though, which is providing free knowledge."[256]

The principle is real, the partnership blurs it, the mission excuses it. Wikipedia Zero ended quietly in February 2018, after India's regulator banned differential pricing and carrier interest faded.[257,258,253] The program itself was defensible, arguably admirable, and its architects called it a stopgap.[256] What outlived it was the argument. A decade later, the same sentence structure would justify selling data services to AI companies. Only the nouns changed.

Building the Product

Lane Becker came in to build Enterprise, the commercial service the Foundation had created the LLC to sell. He was a serial entrepreneur whose startups had been bought by Google and Capital One, with a stint at the civic-tech nonprofit Code for America.[259] He arrived to turn the Delaware shell into a working business.

Enterprise was a premium API, a paid data feed that a company's software could pull from directly. It offered three tiers of access. On-Demand served single articles retrieved as needed. Snapshot served full Wikipedia dumps with periodic updates. Realtime streamed updates as edits happened.[260]

The Structured Contents Initiative turned the raw wiki markup its editors type into machine-readable data. It parsed infoboxes, tables, and references, with quality scoring attached.[261] The content stayed free. The formatting and the infrastructure were the product.

That was the careful line the Foundation drew. Enterprise did not sell content, because anyone could download Wikipedia for free. It sold

bandwidth, speed, format, and service guarantees.[38] Premium access for premium customers.

Launch and Growth

In March 2021, *Wired* broke the story: "Wikipedia Is Finally Asking Big Tech to Pay Up."[245]

On October 25, 2021, Enterprise formally launched from San Francisco.[260] In June 2022, Google and Internet Archive were named as the first customers. Tim Palmer of Google called Wikipedia "a unique and valuable resource."[262]

Revenue grew modestly at first, as shown in Table 13.1.

Fiscal Year	Revenue	Expenses	Net
2022-23	$3.2M	$4.0M	-$0.8M
2023-24	$3.4M	$3.8M	-$0.4M
2024-25	$8.3M	$4.4M	+$3.9M

Table 13.1. Wikimedia Enterprise revenue, 2022–2025[20]

Then came the AI explosion.

AI Pivot

By 2025, AI companies had become the main growth market.[20] Wikipedia reported that 65% of its most resource-intensive traffic came from bots, many of them evading detection by posing as humans.[263]

Jimmy Wales said the strain was real. AI bots had been "absolutely hammering" Wikipedia's servers. "All of the AI bots scraping Wikipedia are actually costing us a lot of money," he told Bloomberg.[264] When the Foundation improved its detection in 2025, it found two things at once. Much of the apparent growth in traffic had been bots masquerading as humans. Genuine human visits had fallen 8 percent year over year.[21]

The Foundation made its pitch on those grounds. Companies that extracted value should contribute value. Enterprise would give the bots a proper feed and a bill.[183]

The growth followed. By January 2026, Enterprise revenue had reached $8.3 million, up 148 percent year over year, and the service was finally profitable.[13,20] New AI-company customers had signed.[19] The contribution problem appeared solved. Companies that had taken now paid.

The March 2024 Admission

But one exchange, in March 2024, complicated the success story. A small AI company building a compliance-first training model asked the Foundation a simple question: was its approach to Wikipedia's licensing correct? The reply, candid and never meant for a press release, conceded that many AI companies probably honored neither the letter nor the spirit of the Creative Commons licenses, and that this one small company was the only one that had ever asked.[18]

The full story of that correspondence, and of what it revealed, belongs to Chapter 16. What matters here is the sequence. The Foundation suspected that companies ignoring the licenses might include its own paying customers.[116] And the partnerships continued.

The compromise was not corruption but dependency. The institution created to protect the shared resource was now profiting from the harvest.

Pragmatism and Its Price

The Foundation's defenders called Enterprise pragmatic. Companies were already scraping Wikipedia.[183] Better to formalize the relationship and get paid for it. Individual access stayed free.[38] Infrastructure cost money, and small donors should not subsidize trillion-dollar firms. "They're not donating in order to subsidize these huge AI companies," Wales told the Associated Press of Wikipedia's donors.[265] A 30% revenue cap, formalized in May 2025, was meant to preserve the nonprofit character. At least 70%

of revenue had to come from donations and grants. Enterprise, at roughly 4% of total revenue, sat well below the line.[20]

Many volunteers saw exploitation instead. A quarter-million volunteers created the content.[13] Enterprise sold access to their labor. They held no share of the revenue and little say over commercial decisions.[38] On Meta-Wiki, the movement's coordination site, contributors put it more sharply: these were not charitable contributions but commercial transactions, the Foundation selling Wikipedia data generated by its volunteer editors. The Foundation's answer was always the same: content was free, infrastructure was not. But the distinction wore thin when "infrastructure" meant formatted access to content volunteers had written for nothing.

There was a deeper question the infrastructure defense dodged. If the only obstacle to free bulk access was cost, the Foundation could afford to remove it. It took in more than $180 million a year and had built cash reserves and a separate endowment besides.[23] A structured, reliable, machine-readable feed of Wikipedia, the very thing Enterprise sold, could have been built once and given to everyone, the way the encyclopedia itself was. That would have served the mission in its plainest form: free knowledge, delivered in the form modern software actually consumes.

Instead the usable infrastructure was reserved for those who paid, while the surplus flowed to a swelling staff, movement grants, and initiatives well beyond the work of maintaining an encyclopedia.[37] The Foundation could have given the machine-readable commons away. It chose to sell it.

Beneath both arguments lay the problem of dependency. Once the Foundation depended on revenue from Google, Amazon, Meta, and Microsoft, enforcement became awkward.[29] You do not sue your paying customers, least of all when you share their worldview, their professional networks, and their vocabulary of "democratizing knowledge." The 30% cap acknowledged the danger, but dependency operates well below any cap. At 4% of revenue, Enterprise was $8.3 million a year. That money

paid for staff, infrastructure, and strategic work.[20] Losing it would hurt. How much non-compliance would the Foundation tolerate to keep it?

WikiPAIDia

In 2002, Edgar Enyedy and a group of Spanish editors forked Wikipedia, breaking off to run their own copy, afraid that advertising would commercialize free knowledge.[57] They were right about the threat. They were wrong about the mechanism.

Commercialization arrived not through banner ads but through enterprise contracts.[38] Not through readers seeing advertisements but through AI companies training on volunteer labor.[18] Not through Bomis, the for-profit company that then ran Wikipedia, taking value, but through the Foundation selling access to firms that might be breaking the licenses that made Wikipedia possible.[44,18]

Enyedy had signed off with "Good luck with your wikiPAIDia" twenty-four years before Wikipedia became, in a sense, exactly that.[57] The revenue model he feared had arrived, just not the way anyone pictured.

The danger was never that commercial interests would come for Wikipedia. That was always going to happen. The surprise was that the institution built to guard against it became the main vehicle for it. The steward had turned into a conduit. What Enterprise built was a pipeline straight to the extractors, and they were already lining up.

Part IV

The AI Reckoning

CHAPTER FOURTEEN

The Extraction Begins

> It would be impossible to train today's leading AI models without using copyrighted materials.
>
> — OpenAI submission to UK House of Lords, January 2024

On May 28, 2020, a research paper appeared on arXiv.org, the preprint server where AI researchers post their work.

Tom Brown and 30 co-authors at OpenAI published "Language Models are Few-Shot Learners," the technical paper introducing GPT-3. Buried in 75 pages of architecture diagrams and benchmark results was Table 2.2, a breakdown of what GPT-3 had consumed.[7]

The table listed five data sources, each measured in tokens—the word-fragments a model reads as its basic units. Common Crawl, scraped from the open web, provided 410 billion tokens, 60% of the training mix by weight. WebText2, a curated collection of outbound links from Reddit, contributed 19 billion. Two collections of digitized books added another 67 billion.[7]

Then came the final line: English Wikipedia. 3 billion tokens. 3 percent of the total.

But the column that mattered was not the one measuring quantity. It was the one measuring quality. OpenAI had assigned each data source an "epoch" weight, a multiplier setting how many times the model would

see each piece of training data. Common Crawl, the bulk of the corpus, received 0.44 epochs. The model would see less than half of it.

Wikipedia received 3.4 epochs. The model trained on Wikipedia more than three times over. The open web got less than half a pass.[7]

The implication was clear. OpenAI considered Wikipedia the most trustworthy source in its training data, by a factor of nearly eight. The encyclopedia that volunteers had spent two decades building had become the quality benchmark against which the internet was measured.[13]

Table 2.2 appeared in a paper read mostly by researchers. No press release celebrated its implications. The volunteers who had created this benchmark were not consulted, compensated, or credited. Their names appeared nowhere in the 75 pages.

The harvesting had begun. Legal scholar James Boyle had a name for the enclosure of shared knowledge: the "second enclosure movement," the expansion of private claims over information and culture once held in common.[266] AI training had found its most valuable target.

Scale of Extraction

GPT-3 was not an outlier. Every major language model that followed would rely on Wikipedia as a quality anchor.

In February 2023, Meta released LLaMA, demonstrating that leading performance was possible with publicly available data. The training data breakdown showed Wikipedia at 4.5%, higher even than GPT-3's 3 percent.[8] The model that proved you could build a frontier AI without proprietary data depended on an encyclopedia built by volunteers.

After LLaMA, transparency evaporated. OpenAI stopped disclosing training data composition.[267] Anthropic never started. Google's Gemini documentation mentioned training data only in the vaguest terms. Disclosure, it had become clear, created liability.

But absence of disclosure did not mean absence of Wikipedia. In January 2024, OpenAI told the UK House of Lords that training leading

Model	Year	Wikipedia Share
GPT-3	2020	3% (3.4 epochs)
LLaMA	2023	4.5%
GPT-4	2023	Undisclosed
Claude	2023-25	Undisclosed
Gemini	2024-25	Undisclosed

Table 14.1. Wikipedia in AI training data

models would be impossible without copyrighted material.[268] Wikipedia, licensed under CC-BY-SA, was exactly such material. The admission applied directly to it.

The Common Crawl shadow made the extraction even larger than disclosed percentages suggested. Common Crawl scraped the open web, and the open web included Wikipedia and its countless mirrors.[269] Mozilla's 2024 analysis found that 64% of 47 major language models used Common Crawl in their training.[270] Wikipedia appeared both directly in training data and indirectly via Common Crawl overlap. The true Wikipedia presence in AI systems exceeded anything Table 2.2 acknowledged.

Citation Dominance

Wikipedia's ingestion showed in what the models produced.

By August 2025, the analytics firm Profound had built a running analysis covering 680 million AI citations. The findings confirmed what the training data suggested: Wikipedia dominated AI knowledge production.[12]

For the world's most popular AI assistant, Wikipedia was not one source among many. It was the foundation of factual knowledge, cited almost as often as the other nine sources in the top ten put together.

Platform	Top Source	Share of Top-10 Citations
ChatGPT	Wikipedia	47.9%
Google AI Overviews	Reddit	21.0%
Perplexity	Reddit	46.7%

Table 14.2. AI citation patterns by platform

Memorization Evidence

The conventional defense of AI training held that models "learned" from data rather than copying it. Training was transformation: raw text became statistical patterns, patterns became capabilities, capabilities generated new text. The original sources were digested, not reproduced.

Nicholas Carlini and colleagues at Google DeepMind proved this narrative false.

The first proof came in 2021, when Carlini's team documented what they called "memorization," the tendency of language models to reproduce training data verbatim when prompted the right way.[9] It was the earliest demonstration, on what is now a tiny model, and everything since has made the finding larger and more systematic.

The 2021 paper, "Extracting Training Data from Large Language Models," showed that training data could be recovered from production models. Larger models were more vulnerable, not less.[9] The capability that made models useful also made them leaky.

In 2023, "Quantifying Memorization Across Neural Language Models" identified three laws. Memorization increases logarithmically with model size. Examples seen multiple times are memorized more. Longer prompts extract more memorized content.[109] That same year, a team led by Milad Nasr demonstrated a "divergence attack" that broke ChatGPT out of chatbot mode and pulled memorized training data out of it in bulk, at roughly a hundred and fifty times the normal rate.[10] This was not a theoretical vulnerability. It was practical extraction from a live commercial product.

Wikipedia was prime memorization territory. High quality meant direct inclusion in training data. Multiple sources meant appearance both in dumps and Common Crawl. Repeated epochs meant GPT-3 saw Wikipedia more than three times over. By Carlini's second law, duplicated data is memorized more, so Wikipedia's quality made it a memorization magnet.[109]

By 2025 and 2026, researchers had turned the finding into hard exhibits: open models reproducing copyrighted novels close to word for word, and the same extraction succeeding against closed production systems from OpenAI, Anthropic, Google, and xAI.[11,271] This was not transformation. It was reproduction at industrial scale.

License Limits

The implications of memorization for CC-BY-SA were devastating.

The license required two things: attribution and share-alike. Attribution meant crediting Wikipedia when using its content. Share-alike meant licensing derivative works under the same terms.[16]

Both requirements collapsed in the AI context.

Attribution. Wikipedia accounted for nearly half of the citations among ChatGPT's ten most-cited sources.[12] But users rarely followed those citations. In Google's AI Overviews, a cited link was clicked on just 1 percent of visits,[272] and chatbot interfaces buried sources even deeper. The attribution existed in the interface but failed in practice. Most users received Wikipedia-derived information without knowing where it came from.

Share-Alike. Did a trained AI model count as a "derivative work"? The early rulings pointed toward no. In *Kadrey v. Meta* (2023), the argument that Llama's model weights, the numerical parameters that make up a trained model, were themselves an infringing derivative work was dismissed as, in the judge's word, "nonsensical."[158]

That dismissal rested on a technical assumption the evidence does not support: that weights cannot contain copies at all. Copyright law defines a copy broadly, and human readability was never the test. United States law defines copies as objects in which a work is fixed and from which it can be "reproduced, or otherwise communicated, either directly or with the aid of a machine or device."[155] Courts had long treated machine-readable encodings, from a program's object code to data loaded into memory, as copies.[273,274]

The memorization research supplied the missing fact. If a model can be prompted to emit a training passage word for word, that passage is encoded in its weights in a form a machine can recover. That is what the statute calls a copy.

The encoding was no accident. Memorization is a design choice, bounded and predictable. A 2025 study measured the ceiling directly: a transformer stores roughly 3.6 bits per parameter, a fixed capacity that training fills.[275] Experiments on Pythia, a suite of open models released for exactly this kind of research, showed which sequences a model would memorize: the pattern scaled with model size and duplication, steadily enough to forecast.[276] For rare facts, the long tail that makes up most of human knowledge, memorization is not a flaw to be tuned away but a mathematical requirement of learning.[277] Weights did not blur their sources into an untraceable haze. They stored them, recoverably, by design.

By the statute's own definition, then, the "nonsensical" label had it backwards. If weights encode even one protected passage, they contain a copy, and whoever distributes a copy of share-alike material owes the same license back. What stopped copyleft was not that the copies had dissolved into the mathematics. It was that no one with standing forced the companies to open the weights and look.

The practical obstacles were real all the same. Kacper Szkalej and Martin Senftleben, in a June 2024 report for the nonprofit Open Future,

found share-alike licenses hard to enforce against AI training. They identified three challenges: tracing uses during training, interplay with fair use exceptions, and identifying protected content in outputs.[69] The license was designed for a world where derivatives announced themselves, where a modified Wikipedia article was still an article, open to inspection. Model weights were not articles. They were billions of numbers whose tie to any one source was buried: recoverable in principle, but slow and costly to prove.

Stallman had created copyleft so that anyone using free software would have to share their improvements. Wikipedia adopted copyleft so that anyone using Wikipedia content would keep their derivatives free.[3] AI companies used that content to build products worth hundreds of billions of dollars. The license that was supposed to protect volunteer labor bound them to share back. No one made them.

The Labor Question

What the extraction meant depended on who was looking.

To Jimmy Wales, AI training validated Wikipedia's achievement. "I'm very happy personally that AI models are training on Wikipedia data because it's human curated," Wales said in January 2026. "I wouldn't really want to use an AI that's trained only on X, you know, like a very angry AI."[24] Wikipedia's mission was to spread free knowledge. AI trained on Wikipedia spread that knowledge faster and wider than the encyclopedia alone ever could. The most sophisticated systems in the world had chosen Wikipedia as their trusted source. By that reading, the volunteers should be proud. They had built the human-curated layer the AI era ran on.

To the volunteers, validation looked different. Volunteer editors had created more than 65 million articles across Wikipedia's language editions.[13] They contributed expecting their work would remain "free as in freedom,"[2] open to all but protected by licensing that would keep deriva-

tives in the shared resource.[16] Then came the extraction, shown in Table 14.3.

Entity	Value Captured
OpenAI (valuation, March 2026)	$852 billion[278,279]
Anthropic (valuation, May 2026)	$965 billion[280,281]
Wikipedia volunteer compensation	$0

Table 14.3. Value captured from Wikipedia training

The companies that trained on Wikipedia built themselves into some of the most valuable enterprises in history, each worth hundreds of billions of dollars. The contributors who created the training data received nothing: not compensation, not attribution in practice, not even the share-alike protection that was supposed to keep derivatives free.[69]

In February 2025, the industry's posture showed plainly. Noam Brown, an OpenAI researcher, tweeted that Deep Research, the company's new automated research agent, "might be the beginning of the end for Wikipedia." He soon deleted the post, though news outlets had already captured it.[282,283] When criticized, his surviving follow-up was candid. Going down this route "essentially means substituting capital for human labor (in this case, Wikipedia editors)."[284] The parenthetical was telling. AI companies did not view Wikipedia as a partner. They viewed it as labor to be automated away.

To the researchers who studied the encyclopedia, the harvesting threatened its survival. Wagner and Jiang, writing in the *Journal of the Association for Information Science and Technology* in 2025, predicted a "vicious cycle."[285] If AI could answer questions without sending users to Wikipedia, why visit? If users did not visit, why edit? If editors stopped editing, content would stale. If content staled, AI outputs would degrade. The system that trained on Wikipedia's quality would undermine the conditions that produced it.

The evidence supported the concern. Human traffic was already declining, down 8% year over year in 2025 once the Foundation's improved bot detection separated real visitors from scrapers.[21,263] The systems Wikipedia had fed were now eating its audience.

None of these readings canceled the others. Wikipedia's quality made it valuable for training. That training absorbed labor without compensation. The scraping might yet erode the encyclopedia's sustainability. The volunteers had created value, AI companies captured it, and the copyleft license meant to protect that work could not function in this context.

Copyleft had always assumed that derivatives would be visible. AI training hid them instead. Model weights were not articles, and training looked nothing like copying. But the copies were there, embedded in the parameters and recoverable by anyone who prompted for them. The copying was real. What failed was the accounting: an industry that preferred not to look, and a steward that never forced it to.

The GPT-3 paper appeared on May 28, 2020. Five years later, Wikipedia topped ChatGPT's citation lists, accounting for nearly half of its most-cited sources.[12] The editors, the arbitrators, the tireless fixers of vandalism saw none of the benefits. No consent, no consultation, no visible credit.

The promise Stallman made, use but share, had become a one-way valve. The using continued. The sharing stopped at the model weights.

CHAPTER FIFTEEN

Selling to the Machine

> Our donors aren't donating to subsidize Sam Altman, they're donating to help Wikipedia.
>
> — Jimmy Wales, Bloomberg, January 2026

The company that used Wikipedia most heavily paid Wikipedia nothing.

OpenAI's ChatGPT cited Wikipedia more than any other source.[12] Its GPT-3 had trained on Wikipedia for 3.4 epochs, more repeat exposure than it gave any other source.[7] The company carried a valuation in the hundreds of billions and annualized revenue near $20 billion.[286] Its payment to the Wikimedia Foundation was zero.[19]

In January 2026, the Foundation named five new paying partners and called the free-rider problem solved—the drain of companies that used the content heavily but paid nothing toward its cost.[19] Enterprise revenue had more than doubled to $8.3 million, profitable for the first time.[20] The company using Wikipedia most heavily was not on the list.

There was nothing mysterious in the choice. Wikipedia's CC-BY-SA license lets anyone obtain the content at no charge, and a company maximizing returns will not pay for access it can get for free.[16] OpenAI behaved exactly as the incentives predicted, which is why the revealing question was never about OpenAI.

It was about the Foundation. If Enterprise existed to enforce the license, the heaviest user's absence would be intolerable. That the Foundation tolerated it showed what Enterprise actually was. Not a compliance mechanism but a voluntary contribution system, revenue from willing payers. The biggest user's absence made the incentive problem visible.

If Enterprise was about fairness, why was the biggest user exempt?

The Market Finds Enterprise

Enterprise had been built to bill a different set of companies. Its product was a premium API, Wikipedia's content in bulk, cleanly formatted and reliably delivered. Launched in 2021, it was aimed at the search engines and voice assistants that had extracted Wikipedia content for a decade, with Google its first paying customer.[38,245] The pitch assumed a slow, steady market of established firms.

Then ChatGPT launched on November 30, 2022, and the market shifted under Enterprise's feet.[287] Within months, language models moved from research curiosity to mainstream product. Every one of them depended on Wikipedia.[7,8]

"The AI companies are here, and they are particularly voracious," Lane Becker, who ran the Enterprise effort, said in an interview with IBM's Think. "We're a general store of knowledge with extremely timely information, and a lot of these organizations are slamming our servers."[288]

AI customers signed, revenue more than doubled in a year, and the service turned profitable for the first time.[19,20] The free-rider problem appeared solved.

The Voluntary Paradox

The declared victory over free-riding concealed a structural problem. Enterprise was entirely voluntary.[38] The license already let anyone copy Wikipedia's pages in bulk with bots, so the paid feed sold convenience, not the content, and no law compelled a company to buy it. But the li-

cense still asked for something in return: attribution and share-alike.[16] An Enterprise contract bought clean data, not compliance with those terms, and the companies that scraped for free ignored them too. Buying the API was optional. Honoring the license was not, though nearly everyone treated it as if it were.

So the companies that saw value in a formal relationship paid. The companies that did not, including the heaviest user, kept scraping without consequence.[183] The numbers made the paradox stark, as Table 15.1 shows.

Metric	Value
Wikipedia share of ChatGPT top-10 citations	nearly half
GPT-3 training: Wikipedia epochs	3.4 (highest)
OpenAI annualized revenue (2025)	$20 billion
OpenAI payment to Wikipedia	$0

Table 15.1. The OpenAI gap: extraction without payment[12,7,286,19]

No AI model leaned on Wikipedia more visibly than ChatGPT, and no company its size gave back less. ChatGPT drew on Wikipedia more than on any other source, whether measured by citations or by training, while OpenAI's revenue climbed toward $20 billion a year and its payments to the Foundation stayed at zero. Money was only part of the debt. Under the license, OpenAI owed attribution and share-alike. It delivered neither, and the Foundation, the one party positioned to enforce those terms, never demanded them.

Wales acknowledged the gap. OpenAI, he said, should "chip in and pay" for the server load its products imposed on the nonprofit.[265] But "should" was not "must." Wales made a moral appeal, not the legal demand the license could have backed, and a moral appeal was easy to ignore. OpenAI ignored it.[19]

Stewardship Against the Heaviest User

The Foundation called Enterprise responsible stewardship.[19] AI companies were already scraping Wikipedia, so engagement and revenue beat a losing fight.[183] The content stayed free while the servers cost money.[263] Becker put the logic plainly: "Wikipedia is a critical component of these tech companies' work that they need to figure out how to support financially."[288] A 30 percent revenue cap held Enterprise near 4 percent of the Foundation's income.[20] That was well below the line where dependency was supposed to begin.

But the stewardship case answered the wrong company. It explained why a willing payer should pay. It said nothing about the heaviest user, who paid nothing and faced nothing. Stewardship that collected from Perplexity while exempting OpenAI was not stewardship of the license.[19] It was a billing arrangement with whoever volunteered to be billed.

Volunteers heard a different argument, and the AI market sharpened it. They had written the content that trained every major model, and Enterprise sold structured access to that labor.[7,8,38] The contributors held no share of the revenue and no vote on the deals.[20] The Foundation replied that the content was free and only the infrastructure was sold.[38] That distinction did not survive contact with what the infrastructure actually was: formatted access to volunteer writing. ChatGPT had turned that writing into the knowledge layer of a $20 billion business.[286,12] The people who produced it were asked to keep producing it for nothing.

From inside the licenses, the gap was a conflict of interest. CC-BY-SA required attribution and share-alike,[16] and the Foundation had already admitted, in March 2024, that many AI companies probably honored neither.[18] Enforcing the license meant suing the violators. Of the likely violators, OpenAI was the one company whose suit would cost the Foundation no revenue, because OpenAI paid none.[19] The companies that did pay might be breaking the same terms. Suing them would end the revenue. So the Foundation sued no one.

That was the equilibrium the structure produced. Move against OpenAI, and the paying customers ask why they were singled out. Move against the payers, and the revenue collapses. Universal non-enforcement was the only stable outcome, dressed as pragmatic adaptation.

An adversarial organization would have built enforcement before taking corporate money. The Foundation never did. Its leadership moved in the same circles as the executives it negotiated with, sharing conferences and a vocabulary of access and scale. Enterprise formalized a relationship the leadership already considered natural. Financial dependency grew from cultural affinity, not the reverse.

Who Paid, Who Used

The January 2026 partner announcement listed who paid. It did not list who used. Google, already a customer, joined the five names added that January, six companies in all that had chosen to formalize their relationship with Wikipedia.[19]

The absences were more telling than the list. Google had almost certainly drawn more from Wikipedia than any AI company, feeding it through two decades of search and now into Gemini, yet Google had at least become a paying customer. The heaviest model-maker, OpenAI, had not, and neither had Anthropic. Both used Wikipedia freely and paid nothing.[19] Neither did xAI, whose Grokipedia had forked Wikipedia, copying it wholesale, in late October 2025.[31]

The more revealing contrast was between the two absentees. OpenAI ignored the license's terms entirely and drew no rebuke beyond Wales's occasional comment.[19,18] xAI, whose fork at least credited Wikipedia and carried its license forward, was about to become the Foundation's loudest grievance.[31,26] The full irony of that pairing belongs to Grokipedia's story, two chapters on. But the direction was already visible: the criticism tracked ideology more than compliance.

The structure explained the silence. Once the Foundation drew profit from AI-company revenue, legal action against AI companies became implausible, whether or not they were the ones paying.[20] The arguments for enforcement already existed. The record held the memorization research showing that models reproduce their training text word for word, the market-harm findings, and the Foundation's own admission that few AI companies honored the license.[9,109,249,18] What was missing was the will, and the structure had removed it.

The press release framed the year as success. "Fair share" was the language, "partnership" the relationship, "sustainability" the goal.[289] The Foundation had built a revenue stream from companies it would not sue, and it knew it.[20,18] It continued regardless.

Enterprise had never been a compliance mechanism. It was something else, and to the volunteers who had written every word it sold access to, it looked a great deal like a racket run in their name.

CHAPTER SIXTEEN

The Capture

> We are monitoring what many LLM companies do with Wikimedia data and generally to be upfront, many may not be compliant with the letter of the Creative Commons rules or the spirit of the licenses.
>
> — Wikimedia Foundation to 273 Ventures, March 2024

In four years of training language models on Wikipedia, only one company asked the Wikimedia Foundation how to comply with the license.[18]

Only one.

273 Ventures was a small legal AI startup, and the model it was building, KL3M, was designed to be trainable without touching anyone's rights: every document licensed, public domain, or otherwise cleared.[1] That discipline would soon make KL3M the first large language model certified by Fairly Trained, a nonprofit that vets AI systems for clean data provenance and documented licensing.[290,291] It also made 273 Ventures the first company to raise the question that OpenAI, Anthropic, Meta, Microsoft, and Google never bothered to ask. How do you comply with CC-BY-SA when training an LLM?[18]

[1] The author was a party to the correspondence described in this chapter. See "A Personal Note," near the end of the book, for a full disclosure of that connection.

The company reached out to the Foundation in early March 2024, trying to do the right thing. KL3M was meant to be an example of responsible AI development, with proper licensing for every piece of training data.

On March 22, 2024, the reply arrived.[18]

The email would never appear in a press release. It would never be mentioned in the Foundation's announcements of AI partnerships. It would never surface in the "fair share" rhetoric that Jimmy Wales deployed when discussing AI companies.[24]

But it would reveal what the Foundation knew, and what it chose to do with that knowledge.

The response was blunt. The Foundation acknowledged that it was watching how AI companies used Wikipedia content. Compliance with Creative Commons terms was, at best, uncertain across the industry.[18]

And then came the admission that exposed an entire industry's willful ignorance. "Their team admitted to us that as of April 2024, we were the only people who had asked how to comply with the licensing requirement when training a model," Jillian Bommarito, the company's co-founder, later reported.[18]

Four years of AI development. Hundreds of billions of dollars in valuation. The largest knowledge extraction in human history. And only one company asked how to do it legally.

The rest did not ask because they did not want to know. The Foundation did not tell them because telling would have complicated the commercial relationships it was building.

The question was not whether the license had failed. The question was whether the steward had abandoned enforcement.

What the Email Revealed

The March 2024 email contained more than an admission. It contained a roadmap of what compliance would actually require.[18]

Attribution, the Foundation explained, was not satisfied by crediting Wikipedia in internal documentation. The license required attribution to reach everyone the content was shared with, not just paying customers. A user receiving an AI-generated answer had to know that Wikipedia was the source. Not in principle, but in practice.[18] The attribution must be "reasonable to the medium."[16]

This requirement was devastating. ChatGPT drew on Wikipedia more than any other source,[12] yet its citations were clicked on barely 1 percent of visits.[272] If attribution was invisible in practice, the "BY" in CC-BY-SA was not being honored.

Share-alike created even deeper problems. The license required derivative works to carry the same terms.[16] If an AI model was a derivative work, its outputs would need to be licensed under CC-BY-SA. OpenAI's proprietary terms would violate the license. So would Google's, Amazon's, Meta's, and Microsoft's.[69]

The Foundation's legal team declined to provide a public letter clarifying these requirements. Instead, they offered an alternative: "complexity is high, but ultimately not-a-barrier" if companies signed up for Enterprise.[18]

The implication was clear. The Foundation knew compliance was difficult. It knew most companies were not achieving it. Its response was not enforcement but sales.

Enterprise became the path of least resistance. Pay for API access. Get structured data. Do not ask about compliance, and the Foundation will not either.[38]

Silence That Followed

Between March 2024 and January 2026, the Foundation took no public enforcement action against any AI company.[136]

It signed five more Enterprise deals and announced them, with some ceremony, as the encyclopedia turned twenty-five.[19] Revenue more than doubled.[20]

It did not mention that most of its customers might not be compliant. It did not explain why compliance was not a condition of Enterprise access. It did not address the gap between what CC-BY-SA required and what AI companies were doing.[19,18]

The silence was itself a statement. The Foundation had chosen accommodation over enforcement. The choice was made deliberately, with full knowledge of what was being accommodated. It would profit from the violations instead of preventing them.

The Grounds They Had

The choice to accommodate is damning only if enforcement was possible. It was. The Foundation had grounds, and more than it ever admitted.

One opening was technical. Researchers at Google DeepMind had shown that language models memorize and reproduce their training data, sometimes word for word.[10,109] Wikipedia, weighted more heavily than any other source in the models that trained on it, was among the likeliest content for a model to retain.[7] If a model reproduced content verbatim, the industry's "transformative use" defense weakened. Transformation implies change. Verbatim reproduction is not change.

Another was economic. As AI products answered the questions people once brought to Wikipedia, its traffic measurably fell.[21] The U.S. Copyright Office had named market harm the single most important factor in fair use, and here the harm was documented, not hypothetical.[160]

The license supplied the rest. CC-BY-SA carried two requirements the companies were not meeting.[16] Attribution was not reaching users in any sense the click-through numbers could dignify. Share-alike was not being honored, because model weights and outputs were proprietary.[69] The "BY" and the "SA" both had claims.

None of this was speculative. Memorization was peer-reviewed.[109] The traffic decline was measured.[21] The first AI training ruling had broken

toward the content owner.[292] The full case for enforcement, factor by factor, was stronger than the Foundation ever conceded.

The Foundation had the grounds. It chose Enterprise revenue instead.

The Precedent: Capture Before AI

Before trillion-dollar companies captured the Foundation, a convicted sex offender proved it could be done.

In December 2011, Jeffrey Epstein's in-house publicity director, Christina Galbraith, sent him an email assessing their options for online reputation management. Wikipedia, she explained, was "a tough nut to crack." On the surface, she wrote, it was "controlled by a morass of copyediting geeks who have nothing better to do than to discuss reference tags."[293]

The contempt was misplaced. The geeks would eventually win. But not before Epstein was, as the *Wikipedia Signpost*, the community's newspaper, later concluded, "partially successful for a decade,"[294] and not before leaked emails revealed exactly how it was done.

The emails, released by the House Oversight Committee in November 2025,[295] documented a systematic campaign to whitewash Epstein's Wikipedia article after his 2008 conviction for soliciting a minor for prostitution. The effort began in 2010 with Al Seckel, a collector of optical illusions[296] who reached Epstein through his partner Isabel Maxwell, sister of Ghislaine Maxwell.[293]

Seckel's December 2010 report to Epstein was triumphant: "Wikipedia was an important victory, as it will always be at the top of the search engine results. Now the headlines do not mention convicted sex offender or pedophile. Instead, Philanthropic work, Epstein Foundation, Promotion of Scientists."[295]

The tactics were documented in his emails. Seckel claimed to have "hacked the site" to replace Epstein's mugshot with a friendlier image. He pushed negative content "to the bottom." He identified the IP addresses of

opposing editors and used them to get the accounts blocked. He removed "toxic suggested search engine terms" from autocomplete.[297]

"Your wiki entry now is pretty tame," Seckel wrote. "Careful editing and wording has muted the effect immensely."[297]

Wikipedia's volunteer defenders noticed. In January 2012, one account, Stgeorge12, was blocked[298] after leaving an edit summary that would become infamous: "I have been asked by Jeffrey Epstein to describe his biography in a professional and accurate way, that does not involve any scandals or disreputable content."[294]

The volunteers blocked accounts. The accounts multiplied. New ones appeared, among them an account called Turvill, alongside a rotation of anonymous IPs.[298] The edit war ran for years.[294]

The real-world consequences were documented. In March 2013, MIT Media Lab director Joi Ito considered accepting a donation from Epstein. A staff member warned him: "You should read his Wikipedia bio, there may be some other things to consider." MIT's later report noted that Wikipedia's article "could have warned Ito," but "also included statements that could be read as undercutting the strength of some of the allegations." The whitewashing had worked well enough to enable institutional rationalization.[293]

The volunteers eventually prevailed. Turvill was blocked in November 2019 after the *New York Times* investigation. The talk page now carries permanent warnings: "Connected contributor: Turvill." "Connected contributor: 71.165.127.242."[298]

But "eventually" was ten years.[294] A convicted sex offender, spending perhaps $10,000 per month on reputation management,[293] had captured his Wikipedia article for a decade.

That decade said as much about an institutional gap as about one determined individual. Wikipedia's only defense was volunteers reverting edits in their spare time. The Foundation provided no escalation path, no funded investigative capacity, nothing built to meet a sustained, well-

financed campaign. The weakness was structural, not a failure of the volunteers' resolve.[294]

The lesson for what followed was clear. If one wealthy individual could overwhelm volunteer defenses for ten years, what chance did "copyediting geeks" have against companies worth trillions?

What followed proved the point on a scale Galbraith never imagined. The companies worth trillions did not need to hack the site or block opposing editors. They had something Epstein's people never did: the Foundation's cooperation. The capture Galbraith described was real. It just happened differently than she pictured. Not one article captured by one client, but the entire commons captured by an industry, with the steward's complicity.

The Economics of Accommodation

Defenders of the Foundation made a practical case worth taking seriously. Enforcement, they argued, was a losing battle. AI companies would train on Wikipedia regardless of legal threats. The content was freely available. The training had already happened. Litigation would be expensive, uncertain, and possibly counterproductive.

Better to engage. Better to collect revenue. Better to ensure sustainability than to die on a principle. The companies were going to use Wikipedia regardless, the Enterprise team argued, so at least some of them should pay something.[288] The 30 percent revenue cap acknowledged the risks of commercial dependence. The Foundation was being careful, working a difficult transition with imperfect options.[38]

The structure, though, looked much like what economists call capture.

The economist George Stigler argued that agencies meant to regulate an industry often end up serving it instead. The mechanism was structural. Concentrated interests, a few large companies with massive stakes, organize effectively. Diffuse interests, millions of affected individuals, do not.[299]

The Wikimedia Foundation was not a regulatory agency. But it was the steward of the copyleft license,[136] and its enforcement decisions affected a quarter million volunteer editors. Those editors had diffuse interests. Each individual's stake was small. AI companies had concentrated interests. Each company's valuation depended on continued access.

The Foundation fit the pattern. It spoke of "partners" rather than "defendants," building cultural alignment with the industry.[19] Financial dependence on Enterprise revenue, paired with the admission followed by no action, completed the dynamic.[20] The capture required no corruption and no bad faith. It required only that the Foundation's financial interests align with the extractors'. Once Enterprise revenue mattered, enforcement became economically awkward.

But the conflict was not purely financial. Foundation leadership had never viewed these companies as adversaries. Federal Election Commission records tell part of the story. Of roughly $295,000 in identifiable political contributions by Foundation employees between 2008 and 2026, more than 99 percent went to Democratic candidates and committees.[22] Just $151, from three donors across the entire eighteen-year span, went to Republicans. The pattern matched Big Tech's executive class almost exactly.

The professional networks overlapped too. Executives with Council on Foreign Relations affiliations, World Economic Forum credentials, McKinsey backgrounds: the Foundation's leadership pipeline drew from the same pools as tech company C-suites.[300,301,302] San Francisco's professional culture shaped both. When tech executives walked into Foundation offices, they met people who spoke their language, shared their politics, attended the same conferences.

The Foundation never developed antibodies against entanglement because it never perceived tech companies as threats. No internal faction asked whether "partners" might be adversaries in licensing terms. No

dissenting voice questioned selling to companies whose compliance was doubtful.

This is enclosure. Not the old closure of common land by English landlords, but the closure of a digital commons by its own guardian, through financial dependence on the extractors. The commons was not locked away. It remained open. But it had lost its only defender, and without defense, openness became vulnerability.

The Opposition That Never Formed

The Linux world had built institutions against exactly this danger: a steward too entangled to enforce. Its enforcement bodies, the Software Freedom Conservancy and the Software Freedom Law Center, sat apart from the corporate money that flowed to the Linux Foundation, so when companies violated the GPL, someone sued.[146]

The Wikimedia Foundation had no equivalent enforcement constituency. Political homogeneity had eliminated the internal voices that might have argued for adversarial enforcement. Nearly every donating employee gave to the same party, and the leadership pipeline drew from the industry's own pools. Partnership felt natural because, culturally, it was. Cultural affinity preceded financial entanglement. By the time Enterprise revenue created dependence, the Foundation had already decided these companies were allies.

That homogeneity had a cost beyond licensing. A culture that generated no internal argument for enforcement also tolerated little dissent of any kind. The community demonstrated as much in June 2026, when it blocked the co-founder indefinitely.[30] Larry Sanger had spent years arguing that Wikipedia's culture was politically lopsided[33] and needed an organized counterweight.[303] When he tried to build one, the machinery that never moved against the extractors moved briskly against him.[30] A community that would not abide a loyal opposition on politics was not go-

ing to grow one on licensing. The monoculture that disarmed enforcement and the monoculture that expelled its co-founder were the same.

The OpenAI Paradox

For the volunteers, the March 2024 admission was a betrayal. They had contributed under CC-BY-SA expecting the license to mean something. It promised that derivative works would stay free, that attribution would follow their words, that share-alike would hold.[16] The Foundation was the guardian of that promise.[136] It admitted the promise was being broken, and it kept serving those who broke it.[18]

The distribution of value was stark. Volunteer editors created the content and received nothing. WMF staff ran operations and drew salaries.[213] Wikimedia Enterprise packaged the APIs and collected millions.[20] AI companies performed the extraction and captured immense valuations. The people who made the value got no compensation, no consultation, and not even the licensing protection they had been promised.

The clearest test of the Foundation's "fair share" rhetoric was its heaviest user.

OpenAI paid Wikipedia nothing.[19]

Asked in early 2026 whether the Foundation would sue the AI companies, Wales answered: "Not yet. Not yet." Conversations, he added, were ongoing.[24]

Not yet. Two words that acknowledged what could be done. Two words that did not say when, or whether, it would be.

But "not yet" was less a personal choice than a structural one. Any leader of the Foundation faced the same pull. A suit against the AI companies would have put at risk the Enterprise revenue the organization had come to depend on. Wales was voicing the logic of capture, not the preference of one man.

If the Foundation would not sue OpenAI, the company using Wikipedia most heavily and paying nothing,[12,19] it would not sue anyone. The legal arguments existed. The will to use them did not.

The bargain at the root of all this was simple. Use my work, but share what you build from it. Richard Stallman had written that condition into copyleft so that no one could take a community's work and wall it off.[3,107] The terms would follow the work wherever it traveled.[16] The commons would protect itself.

The bargain depended on someone willing to enforce it. For Wikipedia, that someone was the Wikimedia Foundation, the institution created to defend the commons.[100] It admitted, in writing, to a company trying to comply, that many AI companies were probably not meeting the terms.[18] The same accommodation continued afterward.[19] Non-enforcement left the commons open to extraction of every kind, the invisible sort that paid nothing and the visible sort that copied everything.

The copyleft promise did not die because the license failed. It died because the enforcer had been captured. The commons stayed open in theory and undefended in fact.

CHAPTER SEVENTEEN

Grokipedia

> The guardrails are off.
>
> — Harold Triedman, Cornell Tech, November 2025

On the morning of October 28, 2025, Harold Triedman opened his laptop and began scrolling through a new website.

Triedman was a graduate student in computer science at Cornell Tech, with an unusual background: before returning to academia, he had spent years as a senior privacy engineer at the Wikimedia Foundation. He knew how Wikipedia worked from the inside—the editorial processes, the reliability standards, the community governance that made the encyclopedia trustworthy.[26]

What he found that morning looked like Wikipedia. It had the same format: articles, infoboxes, citations. It used the same language: encyclopedic, neutral, authoritative. It even acknowledged its debt: small text at the bottom said "The content is adapted from Wikipedia, licensed under Creative Commons Attribution-ShareAlike 4.0 License."[304,25]

But something was wrong.

Triedman clicked on the article for *American History X*, the 1998 film about a reformed neo-Nazi. The article cited viewer reactions to the film. The citations were to Stormfront, one of the oldest and largest neo-Nazi forums on the internet.

Six citations to Stormfront. In an article about a movie criticizing neo-Nazism.[26]

Triedman kept clicking. The citations kept appearing. Not just Stormfront. The conspiracy site InfoWars. The white-nationalist VDare. Sources Wikipedia had blacklisted precisely because they were unreliable, biased, or hateful.[32,305]

This was Grokipedia. Elon Musk's "truthful alternative" to what he called "Wokepedia." The product that had launched the day before to enormous publicity.[304,306]

It was Wikipedia. With neo-Nazi corrections.

October 27, 2025

The day before, Elon Musk had announced Grokipedia on his platform X. The product would "exceed Wikipedia by several orders of magnitude in breadth, depth and accuracy." It would combat the alleged bias of the mainstream encyclopedia. It would offer truth to counter "woke" distortion.[307]

The site launched with 885,279 articles labeled "v 0.1." It crashed within hours from traffic. When it came back, analysts began examining what it actually contained.[308,306]

The answer was Wikipedia.

Forbes flagged the AMD, Lamborghini, and PlayStation 5 articles as copied.[309] The Verge found entries like MacBook Air "almost identical—word-for-word, line-for-line."[310] The plagiarism-detection service Copyscape found 96% similarity between Grokipedia's "Monday" article and Wikipedia's.[311]

On October 31, Musk explained the method on the *All-In Podcast*. The team had Grok, the AI model built by Musk's company xAI, work through Wikipedia's top one million articles: "add, modify and delete." Then Grok would "research the rest of the internet, whatever is publicly available,

and correct the Wikipedia articles, fix mistakes, but also add a lot more context."[307]

The process was explicit. Take Wikipedia. Run it through AI. Publish the result. One million articles, extracted and "corrected" without human review.

Grokipedia was not a Wikipedia competitor. It was Wikipedia with AI modifications.

The Citation Audit

Triedman did not stop at scrolling. He and Alexios Mantzarlis ran a systematic analysis. Mantzarlis directed Cornell's Security, Trust, and Safety Initiative and had led the International Fact-Checking Network.[26,32]

Neither author came to the question from neutral ground. Triedman had worked for the Foundation whose encyclopedia Grokipedia was built to displace, and Mantzarlis came from the fact-checking establishment. A skeptic could fairly discount their judgments. The counts were another matter: specific, methodical, and checkable, and xAI never disputed them.[26]

They scraped 883,858 Grokipedia articles between October 28 and 30, 2025, a full 99.8% of the corpus. They measured its citations against Wikipedia's reliability standards. They quantified what the AI "corrections" had produced.[32]

The audit produced specific counts:

The last two rows isolate the articles Grok wrote on its own, the ones not copied from Wikipedia and so not carried under its Creative Commons license. Those original entries cited unreliable and blacklisted sources far more heavily than Wikipedia did.

Wikipedia had editorial policies to prevent exactly this. Reliable-sources guidelines. A blacklist of sites deemed categorically unreliable. Community review of contested citations. Human judgment about what counted as evidence.[82,305]

Metric	Count
Stormfront citations	42
InfoWars citations	34
VDare citations	107
Total "low-credibility" citations	12,522
Non-CC articles citing "generally unreliable" sources	3.2× more than WP
Non-CC articles citing Wikipedia-blacklisted sources	13× more than WP

Table 17.1. Grokipedia citation analysis (Cornell study)[32]

Grok had no such safeguards. It compiled Wikipedia, then "improved" it by adding sources Wikipedia's editors had specifically excluded.[32]

The Hitler article showed what that meant. NBC News counted 13,000 words before Grokipedia's entry named the Holocaust. Wikipedia's article mentioned it in the first paragraph, with the atrocities front and center.[26]

Grokipedia's original opening read: "Adolf Hitler was the Austrian-born Fuhrer of Germany from 1933 to 1945." The honorific came before the genocide. The structure emphasized economic achievements before mass murder. Not through explicit denial, but through arrangement: what came first, what came last, what received emphasis.[312]

The Intercept put it this way: "No longer do you need a cadre of bureaucrats or the Heritage Foundation to rewrite history books; a metric ton of processing power to help launder ideology through the aesthetics of objectivity suffices."[312]

A More Compliant Fork

The footer on every Grokipedia page named its source: content "adapted from Wikipedia, licensed under Creative Commons Attribution-

ShareAlike 4.0 License." Critics rarely acknowledged what that line represented. It was more visible compliance than ChatGPT offered.[25]

Dimension	Grokipedia	ChatGPT
Attribution visible	Yes (footer)	No
Free access	Yes	Limited (paid tiers $20–200/mo)
Wikipedia-derived content license	CC-BY-SA	Proprietary
Content checkable	Yes	No (black box)

Table 17.2. Licensing compliance comparison[25,34,313]

Grokipedia used two licenses. Wikipedia-derived content stayed under CC-BY-SA, as required. The AI-generated additions carried a proprietary license of xAI's own making, the "X Community License," not the share-alike terms Wikipedia required.[34,16] That structure arguably violated share-alike for the additions. The Wikipedia-derived content itself remained properly licensed.

Plagiarism Today called it "an excellent example of doing the bare minimum to complete the terms."[25] The bare minimum was still more than ChatGPT did. OpenAI absorbed Wikipedia into proprietary model weights, gave end users no attribution, and charged $20 to $200 a month for access.[313] Grokipedia, whatever its content failures, kept its text public, its attribution visible, and its Wikipedia-derived material under CC-BY-SA. Musk did what OpenAI, Meta, Google, and Amazon had done for years, only with the attribution the others withheld.

Asked to comment on the Cornell study, xAI did not respond. Its media inbox returned a three-word autoreply: "Legacy Media Lies."[26] It certainly was not the best way to answer for neo-Nazi citations, but it was arguably less evasive than the flowery corporate nonsense the other AI companies offered when their own use was questioned.

The Foundation Objects

The Foundation objected. Selena Deckelmann, its Chief Product and Technology Officer, told reporters:

"Even Grokipedia needs Wikipedia to exist. Wikipedia's content is open source by design; we expect it will be used in good faith to educate. This issue is especially urgent as platforms like Grokipedia increasingly draw on our articles, selectively extracting content—written by thousands of volunteers—and filtering it through opaque and unaccountable algorithms."[307]

The statement was accurate. It was not enforcement. "Expect good faith" was not a legal standard. "Urgent" was not an action.

Wales was dismissive. "The LLMs he is using to write it are going to make massive errors," he told CNBC at a summit on October 28. "We know ChatGPT and all the other LLMs are not good enough to write wiki entries."[314]

On the charge that Wikipedia was "woke," Wales was characteristically pointed: "We don't treat random crackpots the same as *The New England Journal of Medicine* and that doesn't make us woke. It's a paradox. We are so radical we quote *The New York Times*."[314]

The disdain was justified. But disdain was not enforcement. Wales could mock Grokipedia's quality while the Foundation took no action to stop it.

That restraint was structural. For years the Foundation had signed Enterprise deals with AI companies whose license compliance was questionable at best. It had welcomed AI use of Wikipedia, provided payment came with it. Now it was entangled with the very companies it might otherwise have challenged.[315,183,19]

Singling out Musk when Google, Amazon, Meta, and Microsoft did the same was not principle applied consistently. The contrast made the outrage look selective: the Foundation's sharpest condemnation fell on the one extractor outside its circle of paying partners.[29]

The Foundation did not threaten legal action. It did not assert licensing violations. It did not demand attribution or call for share-alike enforcement. Words were all the steward was prepared to use. Years of inaction had set the precedent.[307]

The Freedom to Fork

One voice complicated the response to Grokipedia. Larry Sanger, Wikipedia's estranged co-founder, had criticized the encyclopedia for years, arguing it had tilted toward mainstream establishment views.[41,316] When Grokipedia appeared, he was cautiously supportive: "The more encyclopedias in the world there are, the happier I am."[317]

His first read was measured: "My initial impression, looking at my own article and poking around here and there, is that Grokipedia is very OK. The jury's still out as to whether it's actually better than Wikipedia. But at this point I would have to say 'maybe!'"[318]

Sanger was not blind to the failures. He called the AI's confident invention of false information "bullshittery."[88] He graded his own Grokipedia entry a "C," "not at all a failure" but with "many bad points." He noted bias running in more than one direction, including to the right.[87] He judged Grokipedia as he would any encyclopedia, by content quality rather than political valence.

His view rested on a principle the copyleft movement had always held: the freedom to fork is fundamental. The Spanish Wikipedia community had forked in 2002 over commercialization fears.[57] Sanger's own Citizendium had forked in 2006 over expertise and governance.[4] The license permitted these forks by design.[16] When governance failed, the remedy was to build something else.

The open question was whether Grokipedia's content failures disqualified it as a legitimate fork. Sanger thought the jury was still out. The neo-Nazi citations and Holocaust minimization suggested otherwise. But

those were editorial failures, the kind better curation could correct. They were not structural violations of what copyleft protected.

Within months, that principle cut against Sanger himself. In June 2026 Wikipedia's community blocked his account indefinitely, after a dispute over his campaign for a reform project and his off-site recruiting for it.[30] He had argued for decades that failed governance justified building an alternative.[55] The encyclopedia he named applied his own logic to him, leaving the freedom to fork as the only freedom he had left.

The Loop Begins

Grokipedia did not stay contained. By January 2026 it had grown to more than 5.6 million articles, approaching the scale of the English Wikipedia. Grok generated new content daily.[319,34]

It also improved, unevenly. A version 0.2 added a correction pipeline: readers could suggest fixes, and Grok reviewed and applied the ones it accepted.[34] Reviewers kept finding fabricated claims and invented citations,[307] and in January 2026 ChatGPT was still surfacing debunked Grokipedia material.[28] But the launch corpus's most notorious failures were eventually scrubbed. By mid-2026, the American History X article that had introduced Grokipedia to the world no longer cited Stormfront at all.[35]

The cleanup, though, arrived after the copying. Wikipedia had trained the AI. The AI had built Grokipedia. The next models trained on a web that had carried the launch corpus, and the contamination reached users with no sign of its origin.[320]

Grok itself, asked whether Grokipedia was trustworthy, gave a remarkable answer:

"Grokipedia is not a fair and unbiased source of information—at least not in the neutral, encyclopedic sense that Wikipedia aspires to. It's a critique wearing encyclopedia clothing."[321]

Of course, with the right prompt an LLM can be steered to say almost anything, and that malleability is itself the problem. Grok could produce that verdict on demand, and xAI published the product anyway: the neo-Nazi sources, the white nationalist references, the Holocaust buried deep in the Hitler article.[26] The guardrails were off. The Foundation's years of accommodation had cleared the path.

CHAPTER EIGHTEEN

The Degradation Loop

> Indiscriminate use of model-generated content in training causes irreversible defects in the resulting models.
>
> — Shumailov et al., Nature, 2024

On January 24, 2026, *The Guardian* ran a story about Sir Richard Evans.[28] He read about himself that morning.

Evans was Regius Professor of History Emeritus at Cambridge, one of the world's foremost historians of Nazi Germany.[322] He had spent years preparing for the Irving trial, the libel case heard in 2000, in which Holocaust denier David Irving sued Penguin Books and historian Deborah Lipstadt for calling him a falsifier of history.[323] Evans's 740-page expert report documented every lie Irving had told, every source he had twisted, every conclusion he had fabricated.[324] The High Court agreed. Irving was a Holocaust denier who manipulated the historical record.[323]

That morning, *The Guardian* reported that ChatGPT was spreading false information about Evans. The source: Grokipedia.[28]

In tests of GPT-5.2, released the previous month, ChatGPT had cited Grokipedia nine times across more than a dozen queries on controversial topics. Asked about Evans specifically, it produced what the article called "debunked information," claims about the historian that contradicted the documented record.[28]

Evans had dedicated his career to documenting how propagandists falsify history. Now an AI system was falsifying information about him. It drew on a Wikipedia copy that minimized the Holocaust, and spread the result to millions of users.[26,312]

The historian who had proved Irving was a liar found himself the subject of AI-generated lies. The mechanism was not a courtroom this time. It was software.

The Test

The Guardian's test mattered less for the count than for where the citations appeared. GPT-5.2 had been released on December 11, 2025.[325] The Grokipedia citations were not scattered at random. They clustered on more obscure topics, exactly where Wikipedia's human curation would have mattered most: contested facts, biographical details, historical interpretations. The AI had learned that Grokipedia was a source, and it cited it where expertise was scarcest.[28]

OpenAI's response was practiced. ChatGPT "aims to draw from a broad range of publicly available sources and viewpoints." Safety filters were applied to reduce "high-severity harms."[28]

The statement acknowledged nothing. "Broad range of sources" included neo-Nazi forums. "Viewpoints" included Holocaust minimization. "High-severity harms" did not include spreading misinformation sourced from Stormfront.

xAI again sent its stock media autoreply: "Legacy media lies." Anthropic, whose Claude also cited Grokipedia for some queries on oil production and Scottish beer, offered no public comment.[28]

The companies landed in roughly the same place: AI systems cite sources, and source quality is someone else's problem. No company committed to excluding Grokipedia from future training, acknowledged the pollution pathway, or engaged with what it meant that ChatGPT now cited an encyclopedia built on neo-Nazi sources.[28,26]

When Models Eat Their Own Output

The contamination was not an accident. It was predicted.

In July 2024, Ilia Shumailov and colleagues at Oxford, Cambridge, and other universities published "AI models collapse when trained on recursively generated data" in *Nature*.[326] The paper documented what happens when AI systems train on AI-generated content.

The findings were systematic. When language models train on their own outputs, quality falls in predictable ways. "Tails" disappear first: the rare, nuanced, carefully curated content that separates high-quality sources from low-quality ones.[326] What remains grows more generic and more homogeneous. Each training cycle compounded the errors of the previous one. Small deviations became large. Rare facts became rarer, then vanished.

Richard Baraniuk's group at Rice University gave the phenomenon a vivid name: Model Autophagy Disorder, or MAD. The culprit was synthetic data, text or images generated by AI rather than by people. "The problems arise when this synthetic data training is, inevitably, repeated," Baraniuk explained, "forming a kind of a feedback loop—what we call an autophagous or 'self-consuming' loop."[327]

The team documented visible effects in image models: "gridlike scars," generative artifacts that compounded with each cycle.[327] In text models the effects were subtler. Lexical diversity narrowed. Syntactic variety thinned. The semantic richness that made language meaningful drained away.

The early alarm needs a caveat. Those experiments assumed each model's output *replaced* the human data before it, an extreme case, and later work complicated the picture. When synthetic data instead accumulates alongside real data, or is filtered and verified before reuse, the degradation can level off rather than compound.[328] Whether the loop rots a model turns out to depend on how the synthetic data is made and managed. That is a warning, not a reprieve. The danger is not AI-generated

content as such, but *uncurated* AI content fed back in without checks, which is exactly what an open web of unlabeled machine text supplies.

The Synthetic Web

The research predicted pollution. The data confirmed it was already happening.

Metric	Value	Source
New webpages containing AI-generated content (April 2025)	74.2%	Ahrefs[329]
AI-written pages in top-20 Google results	11% to 20%	Originality.ai[330]
AI "news" sites (since 2023)	49 to 1,271	NewsGuard[331]
Projected AI-generated online content by 2026	up to 90%	Europol[332]
Hallucination rate, one frontier model (benchmark)	79%	OpenAI[333]

Table 18.1. AI-generated content contamination metrics

Two of those numbers were estimates rather than measurements: the 90 percent was a forecast, and the 79 percent came from one model on a hard factual benchmark. The direction they pointed to was not in dispute. By 2025, most new web content carried the mark of AI, and Wikipedia's seven million human-curated English articles were an island in an ocean of that content.[93] That ocean was feeding back into the AI systems that had created it.

What the Collapse Research Missed

The model-collapse papers documented quality degradation. The Grokipedia-ChatGPT loop revealed something the research had not anticipated: ideological contamination.[28]

Model collapse assumed random noise. AI training on AI would produce blander, less diverse outputs. The degradation was statistical, not directional. Errors would average out.

Grokipedia was not random noise. It was systematic ideological manipulation. This was the same Grokipedia whose launch corpus drew on Stormfront, VDare, and InfoWars, and ran 13,000 words on Hitler before naming the Holocaust.[32,26] It was the same encyclopedia in which Richard Evans's scholarship was contradicted by AI-sourced fabrications.[322]

The loop did not just degrade. It laundered. Content that would be flagged at once as extremist gained the veneer of an "encyclopedia" through Grokipedia, then the authority of "the world's most advanced AI" through ChatGPT.[325]

In a separate Intercept report, LK Seiling, an AI researcher at the Weizenbaum Institute in Berlin, identified the mechanism. Grokipedia was "cloaking misinformation." Wikipedia's authority rested on governance structures that were, in Seiling's words, "all visible and auditable."[312] Grokipedia mimicked that authority without the accountability. The loop was not just self-consuming. It was self-corrupting.

From Citogenesis to Industrial Laundering

Wikipedia had long faced a smaller version of the problem, called citogenesis.[334] Someone adds false information to Wikipedia. A journalist copies it. A Wikipedia editor then cites the journalist as a "reliable source," and the fabrication appears verified. The classic case was a fake middle name added to a celebrity's page: wait for news outlets to copy it, then cite those outlets as proof. The lie became "sourced" by passing through another publication.

Citogenesis was a human-scale problem. Editors could catch it. The traces were visible. The loop could be broken.

The Grokipedia-ChatGPT loop was citogenesis run at machine speed. Wikipedia's quarter-million editors produced tens of millions of human-

curated articles.[93] Major AI models trained on Wikipedia, which supplied a few percent of the data behind systems like GPT-3 and LLaMA.[7,8] Grokipedia copied Wikipedia and added neo-Nazi sources.[26] GPT-5.2 then trained on the open web, Grokipedia's articles included, and ChatGPT cited Grokipedia back as an authority.[28] Users received contaminated information carrying no trace of its origin.

The difference from old-fashioned citogenesis was where the traces went. They were buried in model weights. The sources were not cited. They were absorbed. No editor could find where the contamination entered. No correction could remove it from a trained model.[326] The corruption was embedded in the infrastructure.

Fairness requires admitting that the loop did not spare Wikipedia's own flaws, and that Grokipedia's stated grievance had a real target. Wikipedia is not a neutral mirror of the world. Peer-reviewed work has found measurable political slant in its articles, heaviest in the older, lightly-edited ones.[335]

Its handling of contested subjects, from the Israeli-Palestinian conflict to the history of the Holocaust in Poland, has drawn sustained and documented disputes over sourcing and framing.[336,337] One entire language edition, the Croatian, was captured for the better part of a decade by a handful of administrators pushing far-right politics, until the Foundation's own investigation confirmed it.[338]

Those biases and errors were absorbed by the models too, and laundered outward the same way, years before anyone built a fork designed to do it on purpose. Grokipedia made the problem worse and more deliberate. It did not invent it.

Worse, some of it was deliberate. Researchers had a name for intentional poisoning: LLM grooming. The American Sunlight Project, which coined the term, documented the Pravda network in 2025: a Russian influence operation generating 3.6 million articles a year.[339] The content was designed to be absorbed by AI training pipelines, and the project classified

the operation as a new form of "Foreign Information Manipulation and Interference."[340]

Grokipedia was not built by foreign adversaries, but the mechanism was identical. Generate ideologically skewed content at volume, place it where AI systems would ingest it, and watch the distortion propagate. Wikipedia's community had built guardrails over two decades: reliable-sources rules, neutral point of view, and editorial review.[82] None could function here. The poison bypassed them entirely.

Once it was in, it stayed. Watermarking existed but was not widely deployed. Provenance tracking was incomplete. Model weights could not be inspected to show which training examples had shaped which outputs. The irreversibility was the point. Training a fresh model from scratch on verified human content grew exponentially more expensive as synthetic content came to dominate the web. The economics pushed toward using whatever data was available, tainted or not.

A License Left Unenforced

CC-BY-SA's bargain was simple. Credit the source, and license your derivatives under the same terms.[16] It assumed derivatives you could recognize.

AI companies trained on Wikipedia without clear attribution. Model weights were proprietary. AI outputs were proprietary. The Foundation had conceded the likely non-compliance in writing,[17] and no enforcement followed.

Grokipedia, by contrast, copied Wikipedia with visible attribution and kept the Wikipedia-derived text under CC-BY-SA, layering its AI-generated additions under the proprietary "X Community License."[34,25] That two-license structure raised share-alike questions for the additions, but the borrowed content itself was properly licensed.

The 2024 Open Future analysis named the structural problem. Concepts like "adapted material" and "technical modification" were "largely

absent in AI training workflows."[69] Model weights did not announce what they contained. The copies were there, buried, waiting for anyone willing to dig. Surfacing them would have taken an enforcement effort the Foundation never mounted. The degradation loop was the result.

Buried in the Weights

The mechanism that spread lies about Richard Evans was not a courtroom, or a publishing house, or even a website that could be corrected. It was model weights: billions of parameters encoding the lies in a form no editor could correct and no revision could undo.

The loop revealed two kinds of failure, and they were not equal. Grokipedia's content failures were severe: neo-Nazi sources, Holocaust minimization, ideological curation.[26,312] These were legitimate criticisms. Yet on the license terms themselves, Grokipedia was the compliant half of the pair, and ChatGPT's structural failures were worse from a copyleft view: no attribution, proprietary outputs, paywalled access.[17] The Foundation condemned the visible content failures while enabling the invisible structural ones.

The commons Stallman sought to protect was enclosed from two directions.[107] One was visible forks with editorial failures. The other was invisible extraction with structural violations. The latter did more damage. The former drew more criticism.

The degradation loop was not just a technical failure. It was the logical endpoint of non-enforcement, the consequence of a compromised steward. When the institution meant to protect the shared resource became financially dependent on those harvesting it, the resource lost its only defense. The copyleft Stallman built as armor sat unworn.

Enforcement was expensive, but the deeper failure was institutional will: no one inside the institution saw a reason to put it on. Political homogeneity had bred out adversarial instincts. To an institution staffed by people who saw the executives as peers, "partnership" was a default,

not a decision. The defense failed inside the institution meant to provide it.

This was mechanized loss, damage that perpetuated itself without further human intervention. Each training cycle absorbed the errors of the last. Model weights could not be "cleaned."[326] Human editors could correct Wikipedia. They could not correct the weights of systems they had no access to. The harm compounded on its own. The off switch did not exist.

The encyclopedia anyone could edit had trained the AI that trained the AI that now cited its corrupted copy, while the institution that should have protected it collected Enterprise revenue instead.[20]

CHAPTER NINETEEN

The Factions

> Wikipedia's brand is reliability, traceability of changes, and 'anyone can fix it.' AI is the opposite of these things.[24]
>
> — Bawolff, Wikipedia editor, 2025

On January 15, 2026, the Wikimedia Foundation announced new AI partnerships on Wikipedia's twenty-fifth birthday. Amazon. Meta. Microsoft. Perplexity. Mistral AI.[19]

Around the world, four editors read the announcement and reached four conclusions.

In Singapore, Robert Sim, the newly named Wikimedian of the Year (whose two decades of editing Chapter 22 follows), read it as a demographic warning. Tens of thousands of edits had taught him how much the encyclopedia depended on hands like his, and new editor registrations were down 36% since 2016.[341,249] If AI replaced readers, who would ever become an editor?

In the United States, a volunteer from WikiProject AI Cleanup read it as betrayal. His team had spent months tagging thousands of articles for AI contamination and winning hard policy fights to remove it.[342] The Foundation had just signed five more AI deals anyway.[19]

On Meta-Wiki, the movement's coordination site, a contributor read it as a transparency failure, arguing that the partnerships should be max-

imally transparent because the Foundation was selling Wikipedia data generated by its volunteer editors. No vote had authorized the partnerships. No community consensus had been sought.

In San Francisco, Foundation leadership read it as a milestone. Enterprise was profitable for the first time.[20] The "fair share" message was working.[183] Wikipedia would survive the AI era by adapting to it.

Same announcement. Four incompatible readings. The fracture that had been forming for years had finally set.

The Pro-Enterprise Pragmatists

The Pragmatists controlled the institution.

Jimmy Wales made the public case. AI companies should pay their "fair share." They were already scraping Wikipedia, so it was better to formalize the relationship and collect revenue than to fight a losing battle.[183] Wales was personally glad, he said in January 2026, that AI models were training on human-curated Wikipedia rather than only on the angrier corners of the internet.[24]

Lane Becker kept making the business case he had made since the AI wave arrived: the companies were here, they were voracious, and Wikipedia was critical to their work. They needed to figure out how to support it financially.[29]

The evidence the Pragmatists cited was real. Bot traffic strained the servers, accounting for 65% of the most resource-intensive requests.[263] Human traffic was falling, down 8 percent year over year by the Foundation's own bot-filtered count.[21] Enterprise revenue, newly profitable and growing fast, offset those costs. A ceiling formalized in May 2025 held commercial revenue under 30 percent of the total, keeping the nonprofit balance intact.[20]

From this view, Wikipedia's adaptation to AI was working. The Pragmatists were keeping the lights on.

The Copyleft Purists

The Purists answered to the license.

Wikipedia was built on CC-BY-SA, a license promising that derivative works would stay free.[16] Contributors had trusted that promise, and the Foundation was its steward.[136]

The March 2024 admission had exposed the breach: non-compliance suspected, put in writing, and followed by more partnerships.[17,19]

The June 2024 Open Future report confirmed what the Purists feared. Share-alike licenses were largely ineffective when the material was used to train AI models.[69] Without an enforcer, the legal framework Stallman had helped build could not protect the commons in the AI era.[107]

If the license went unenforced, what remained of the promise? Contributors had written more than 65 million articles under terms the Foundation would not uphold.[13] The Meta-Wiki contributor from the anniversary announcement had put it plainly: these were not charitable contributions but commercial transactions in volunteer-generated data.

The AI Skeptics

The Skeptics fought on the articles themselves.

WikiProject AI Cleanup, founded in December 2023,[343] had become the front line. By October 2025, the project had flagged thousands of articles: LLM-generated content with fabricated citations, impossible syntax, telltale phrases.[342] The volunteers built detection methods, documented the patterns,[344] and won community consensus for action. Through 2025 they turned that consensus into policy, prevailing in a string of disputes over how far AI could reach into the encyclopedia.[342]

But the Skeptics knew their victories were partial. The sophisticated AI content slipped through. A Princeton study found that 4.36% of new English Wikipedia articles created in August 2024 contained significant AI-generated content.[345]

The contrast with AI was the heart of their case. Wikipedia's value came from transparency, from every edit being traceable and reversible.[346] AI models were black boxes, their training data invisible and their outputs unattributable.[10]

The Volunteer Advocates

The Volunteer Advocates asked the simplest question: whose labor, whose profit?

A quarter million unpaid editors had created the content that trained every major AI system.[13,7,8] ChatGPT leaned on Wikipedia above everything else it cited.[12] AI companies had built billion-dollar businesses on this labor. The compensation was zero.

The Conversation, writing on Wikipedia's twenty-fifth anniversary, asked whether we were witnessing "a new 'tragedy of the commons', where volunteered knowledge becomes raw material for systems that themselves may produce unreliable material at scale."[347] For the Advocates, the commercial edge of that tragedy cut deepest. The systems profited, and no one had asked the volunteers.

The numbers were stark.

Metric	Value
Active volunteer editors	about 250,000
Wikipedia share of ChatGPT citations	nearly half
Enterprise revenue (FY 2024-25)	$8.3M
Volunteer compensation	$0
New editor registrations (2016-2025)	-36%

Table 19.1. Volunteer labor and value extraction[13,12,20,249]

Edit volume was up 37 percent even as the editor rolls thinned: fewer people doing more work.[348] The people who created the content that powered AI were being asked to work harder for nothing in return.

The Battles That Drew the Lines

The Skeptics' fights were not abstract. Through 2025, disagreement hardened into open conflict.

It became visible in June, with the summaries revolt. On June 2, 2025, the Foundation launched "Simple Article Summaries," AI-generated summaries built on Aya, a model from the AI company Cohere, and offered as an opt-in to 10% of mobile readers. The community had not been consulted.[349]

The response was immediate. "Yuck," one editor wrote. "I sincerely beg you not to test this, on mobile or anywhere else," wrote another. "This would do immediate and irreversible harm to our readers and to our reputation as a decently trustworthy and serious source." Within 24 hours the experiment was paused. No vote had been held. The Foundation had launched, the community had rejected, and the feature had died.[349] Crisis and backlash had replaced governance.

In July came Wales's ChatGPT proposal. Jimmy Wales suggested using ChatGPT to help with Articles for Creation reviews, the process by which new submissions were evaluated for publication.[24,350]

The Skeptics answered with testing. They gave ChatGPT real review tasks, and the results were damning. It could not reliably apply Wikipedia's notability guidelines, and it made errors that experienced reviewers caught at once. By late August the proposal was dead.[24] Wikipedia's founder had learned that his community would not follow him on AI.

August brought the G15 victory, a new rule for deleting the most egregious AI content on sight. Before G15, removing AI-generated content required week-long deletion discussions while the content accumulated. WikiProject AI Cleanup pushed for something faster.[342]

In August 2025 the relevant RFC closed with overwhelming consensus. "I am closing this per SNOW," the closer wrote, shorthand for a debate so one-sided that further discussion was pointless.[351] G15 allowed immediate deletion of articles containing "communication intended for the user," a

phrase such as "Here is your Wikipedia article on…", or fabricated references such as non-existent papers and invalid ISBNs. It was a real win. It was also, as Ilyas Lebleu of the cleanup project admitted, "a band-aid for the most egregious examples of AI-generated submissions, though the larger problem will continue." G15 caught the obvious cases. The subtle contamination remained.[342]

Different Questions

By late 2025, the camps were no longer arguing over one policy. They were defending different definitions of Wikipedia itself: infrastructure to the Pragmatists, a promise to the Purists, human curation to the Skeptics, their own labor to the Advocates.

These were not positions on one spectrum. They were answers to different questions. No consensus mechanism could reconcile them, because the factions did not share premises about what consensus should protect.

The structure made agreement harder still. The Foundation, the affiliates, and the editors had separate governance bodies with no way to make binding joint decisions.[352] The Foundation carried institutional pressures of payroll, infrastructure, and legal exposure that individual editors did not, while those editors absorbed a labor extraction the Foundation benefited from. Enterprise contracts were confidential, so editors could not judge what the Foundation was actually promising AI companies.[38] The licensing questions underneath it all were genuinely unsettled, and neither side could claim clear legal high ground.[69]

One more barrier ran beneath the rest: the monoculture. The employee donations ran more than 99 percent to one party,[22] the professional networks overlapped with the AI executives' own, and the vocabulary matched. Nobody pushed for adversarial enforcement because nobody inside was built to.

The same logic shaped enforcement inside the community. In June 2026, as Chapter 3 recounts, editors invoked WP:BATTLEGROUND, the

rule against treating Wikipedia as a contest to be won,[353] to block co-founder Larry Sanger indefinitely for organizing an openly declared faction, WikiProject Intellectual Diversity, and recruiting for it off-wiki.[30,39] The charge was not baseless: steering an outside audience into internal votes is a real violation.[354]

But the asymmetry was procedural. WP:BATTLEGROUND could only be invoked against a side that had declared itself a faction. The incumbent alignment never had. It needed no WikiProject and published no manifesto. It worked through the ordinary machinery of consensus, in a culture whose measurable political donations ran almost entirely one way, and so it never registered as a faction the rule could reach.[22] The community's own enforcement tools could see only the side organized enough to name itself.

"This permanent imbalance creates constant tension, wastes time and energy, and ultimately weakens our entire movement," Christophe Henner wrote in 2025 of the Foundation-community power structure.[352]

The Override

Anyone who doubted where final authority lay had two recent demonstrations to study.

The first was the Movement Charter. For three years, an elected drafting committee had worked on a constitution for the movement: a document meant to define, at last, how power would be shared among the Foundation, the affiliates, and the volunteers.[355] In the summer of 2024 the movement ratified it. Nearly three-quarters of individual voters approved, 1,710 to 623. Among affiliate organizations, the margin was 84 percent.[355]

The Board of Trustees voted not to ratify, 11 to 1.[356] It took that vote on July 8, 2024, ten days before the community's results were even formally announced.[356,357] The lone dissenter, Mike Peel, held one of the community-selected seats and told the Board he supported the charter.[358]

Three years of drafting and a supermajority ratification produced, in the end, a Board statement offering alternatives.[359]

The second demonstration came in October 2025, during the election for the Board's own community-and-affiliate seats. Five days before voting opened, the Board announced it had removed two of the six candidates from the ballot, both shortlisted by the movement's affiliates. The decision was unanimous, the reasons confidential.[360] The Board had amended its own candidate-review rules days earlier, adding screening for "reputational, financial, operational, or other types of risk" judged on "subjective criteria like a candidate's judgment."[360] The Board chair explained that the Foundation needed "a strongly unified board" to steer through "difficult global headwinds."[360]

A reform petition gathered more than a hundred signatures, and community groups whose candidates had been struck published objections. The vetoes stood.[361,362] One of the removed candidates, Lane Rasberry, drew the conclusion himself, on-wiki: "the board simply does not feel that I am a suitable candidate," he wrote. "Who I am and what I do is simply not what the other board members want to join them."[363]

By late 2025 the Foundation was under siege from outside: congressional letters, litigation abroad, coordinated campaigns against its credibility. The vetoes read less like a power grab than like an institution in a defensive crouch, pruning perceived risk before adversaries could weaponize it.[364] But for the volunteers, the motive mattered less than the mechanism. A community that voted for a charter watched its Board reject the result before it was announced. A movement that shortlisted its candidates watched them struck from the ballot without stated cause. Twice, in writing, the Board had chosen unity over representation.

The loop closed there for the factions. Voice had been tried at every scale, from talk pages to RfCs to a ratified constitutional charter, and the institution had shown, politely and unanimously, that voice did not bind it.

The Fork That Could Not Be Repeated

The impulse to break away was old. In 2002, Spanish editors had forked over advertising plans from Bomis, the for-profit company that then hosted Wikipedia. Edgar Enyedy had led that exodus with a parting jab about a commercialized "wikiPAIDia." The fork failed against Wikipedia's network effects, and the encyclopedia absorbed the loss and kept growing.[57]

Twenty-four years later, the same fears had arrived through a different door. The 2002 editors feared advertising would compromise independence.[57] In 2026, Enterprise revenue made the Foundation dependent on AI companies. They feared commercial interests would introduce bias, and Enterprise served non-compliant customers anyway.[17] They feared volunteer labor would be monetized without consent, and Enterprise revenue now flowed from volunteer content.[20] Enyedy had been right about the threat and wrong only about its timing and shape. Commercialization came not through banner ads but through API access for companies training AI on volunteer work.[19]

This time, though, the exit was closed. Wikipedia's dominance was far more complete than in 2002, its brand synonymous with "encyclopedia." More than 65 million articles across languages could not be easily rebuilt.[13] Any fork would face the same AI extraction, because scraping did not depend on who hosted the content. And organizing a quarter million editors was a different task from organizing a few dozen Spanish contributors.[57]

Forking solved nothing. Staying meant accepting capture. Sanger had tried both, the fork that never scaled and the reform that got him blocked, and ended with neither door open.[78,30]

The Civil War Already Here

The community that celebrated Wikipedia's twenty-fifth birthday[13] was not the community that had launched the encyclopedia in 2001.[47] The founding dream, that radical openness could produce reliable knowledge

and keep it free forever, had collided with a world that extracted value from openness wholesale.

None of the four answers was wholly wrong, and none could convince the others. The conflict was not about policy but about identity, about whether the enclosure of the commons could be undone, resisted, accommodated, or only mourned.

The civil war was not coming. It was already here, fought through policy RFCs, funding appeals, and quiet decisions to contribute less.

Wikipedia at twenty-five: still free for anyone to edit,[346] feeding the models that may make editing obsolete,[7,8] governed by a foundation that sold access to the AI companies,[19] built by volunteers who were never asked.

Part V

What Remains

CHAPTER TWENTY

The Prophecy

> I started crying right there in the equipment room. Seeing the machine there, dead, with nobody left to fix it, it all drove home how completely my community had been destroyed.
>
> — Richard Stallman, recalling the PDP-10's end

Stallman built copyleft out of grief. The dead PDP-10, the colleagues Symbolics hired away, the ultimatum that company delivered on his twenty-ninth birthday: that wound produced a legal weapon. Copyleft would ensure no one could take shared work and refuse to share it back. Use my code, the license said, but release your improvements under the same terms. He spent the rest of his career hardening it.[2]

The weapon outgrew software. The GNU Free Documentation License, released in 2000, carried copyleft from code to text.[107] When Wikipedia launched in January 2001, it adopted that license.[47] Stallman's promise now covered an encyclopedia. Hundreds of thousands of volunteers wrote under it, trusting that their work, and anything built from it, would stay free.[365,5]

Forty years after the birthday betrayal, the betrayal returned. AI companies took Wikipedia's articles, trained proprietary models on them, and shared nothing back.[7,8] The scale was no longer one company and one

codebase. It was a trillion-dollar industry and most of human knowledge.[14] The weapon Stallman had forged for exactly this moment sat unused.

Symbolics and the AI Companies

The parallel between Symbolics and the AI companies is close in the ways that matter here. In 1982 the commons was Lisp Machine code.[2] In 2024 it was the English Wikipedia's 6.7 million articles.[366] The contributors were MIT hackers, then hundreds of thousands of editors.[365] The promise was a gentleman's agreement to share improvements, then the ShareAlike clause of CC-BY-SA.[16] The betrayal was the same in both: take the shared work, lock the improvements away, give nothing back.

Only the response differed. What changed was disposition, not circumstance. Stallman treated corporations as adversaries by default.[2] He assumed betrayal and built his defenses before it came. The Wikimedia Foundation treated the companies extracting its work as peers.[29] When the AI firms arrived to extract, they found colleagues, not adversaries.

The two responses are worth setting side by side. When Symbolics betrayed the community, Stallman cut their link, declared war, and spent two years keeping the lab's version alive without them. Then he built the legal machinery to stop it happening again.[2]

When AI companies took Wikipedia's content, the Foundation disconnected nothing. It declared no war. It built Enterprise and sold faster access to the same companies whose compliance it doubted.[38,29,18]

Enforcement Never Happened

The deeper point is not that copyleft failed. Copyleft was never tested.

A case for enforcement existed. Language models reproduce stretches of their training text verbatim.[17] Wikipedia's traffic was falling as AI answers replaced the clicks that once reached it.[21] And the courts had begun to move. In *Thomson Reuters v. ROSS Intelligence*, a judge held that training an AI on protected material was not fair use when the product

competed in the same market.[292] The evidence was assembling. The argument had a shape.

The March 2024 admission sat in the Foundation's own correspondence: many of its AI customers, it conceded, likely did not comply.[18] Yet it filed no lawsuit. It made no public demand. It published no enforcement position. The institution that should have defended the commons chose to protect its commercial relationships instead.

It took the money.

Pretend Intelligence

On January 23, 2026, Richard Stallman spoke at Georgia Tech.[367] At seventy-two, he was unchanged.[368] He refused to call large language models artificial intelligence and offered his own term: "pretend intelligence."[367]

"I decided not to be part of their campaign to convince people that those things are intelligent," he said. "Every time you call them AI, you are endorsing the claim that they are intelligence and they're not. So, let's refuse to do that."[367]

He demanded that AI output be labeled. He argued that harmful uses should be regulated by law, not licensing. He was, as ever, exact about language and immovable about principle. But he said nothing about Wikipedia. Nothing about copyleft enforcement. Nothing about the line running from Symbolics to ChatGPT. The man who built the weapon never mentioned that it had gone unused.

The Free Software Foundation, which Stallman founded in 1985, had tried.[2] In October 2024 it published criteria for "free machine learning applications," requiring that training data and scripts grant the four freedoms, the rights to run, study, share, and modify.[369] At FOSDEM, a European free-software conference, it went further in early 2025: "We cannot say a ML application 'is free' unless all its training data and the related scripts for processing it respect all users, following the four freedoms."[369,370]

The criteria were admirable. They were also late. By October 2024 the major models had already trained on Wikipedia.[7,8] ChatGPT had more than 200 million weekly users.[371] The weights were proprietary. The extraction was finished. The FSF described what AI should be, not how to enforce against companies that had already taken what they wanted.

The scholars reached a similar verdict. Researchers at Amsterdam's Institute for Information Law, writing through Open Future in June 2024, judged ShareAlike licenses largely ineffective against AI training. The problem, they argued, was structural. License concepts like "adapted material" had no clear meaning inside a training run. Tracing an output back to its source, in their view, was close to impossible.[69]

But "largely ineffective" was not "unenforceable." The same study mapped paths that might have worked: rights reservation, opt-out mechanisms, extensions to the ShareAlike clause.[69] The Foundation pursued none of them.

Would the license have held if someone had tested it? No one knows, because no one tried. The defenders chose not to defend.

The Guards Were Aligned

On March 16, 2024, Stallman turned seventy-one.[2] That same month, the Foundation put its compliance doubts in writing. Nothing followed, and the deals continued.[18,29] The wound and the weapon shared a birthday. The weapon was inherited, and it was never raised.

The prophecy Stallman made was not that clever lawyers would defeat copyleft. He had built the license to survive even untrustworthy stewards. The terms would propagate through every reuse, whatever the steward did.[16] What he could not design around was a steward that did not want to enforce.

The wall Stallman built still stood. The legal mechanism remained available. Enforcement only needed an enforcer with resources, will, and independence. The Foundation had the resources, more than $180 million

in annual revenue.[23] It lacked the will. And it was not independent. Its leadership was cut from the world of the firms taking the work.

The wall was never breached. The evidence points less to a simple purchase than to prior alignment. The guards had already chosen a side.

The question the commons leaves behind is not whether the license could have protected it. The question is why the people charged with defending it chose not to try.

CHAPTER TWENTY-ONE

The Case That Wasn't Made

> The effect of the use upon the potential market for or value of the copyrighted work is undoubtedly the single most important element of fair use.
>
> — U.S. Copyright Office, quoting the Supreme Court, May 2025

The evidence that should have ended the debate over AI training and copyright was already public. As Chapter 14 showed, researchers had already proven that large language models memorize their training data and can be made to reproduce it verbatim. The sharpest demonstration came from Google DeepMind: under an extraction attack that Milad Nasr and Nicholas Carlini ran against ChatGPT, over five percent of its output was direct copying from the training set. Fifty-token sequences came back word for word.[10]

The findings were reproducible. The grounds for action were clear.

And the Wikimedia Foundation, whose volunteers' content topped ChatGPT's citation lists, filed no lawsuit.[12] Made no public statement. Signed more Enterprise deals.[19]

Memorization Is Copying

The AI industry's defense rested on a narrative: models "learn" from data rather than copy it. Training is transformation. Raw text becomes statistical patterns. Patterns become capabilities. Capabilities generate new text. The original sources are digested, not reproduced.[372]

It was a narrative the memorization research had already dismantled: memorization grows with model size and repetition, and extraction works against deployed products, not just laboratory models.[109,10] What mattered for a lawsuit was the evidentiary residue that research left behind:

Study	Finding	Source
Patronus AI	GPT-4 produces copyrighted content 44% of tested prompts[373]	2024
Cornell	ChatGPT retrieved 72 of 240 poems verbatim[374]	Jan 2024
Ahmed et al.	Near-perfect Harry Potter reconstruction[271]	Jan 2026

Table 21.1. AI memorization studies

The legal significance was direct. Copyright infringement requires proving copying occurred. Verbatim reproduction is the clearest form of copying.[160] If models reproduce inputs, the transformative defense becomes much harder to sustain.

Wikipedia was GPT-3's most-repeated training source, cycled through more often than anything else in the corpus.[7] The quality that made it valuable for training also made it likely to be reproduced.

Proving copying is only the first step. U.S. copyright law lets someone reproduce protected material without permission when the use qualifies as "fair use." Courts weigh that defense through four factors written into the Copyright Act: the purpose and character of the use, the nature of the work, the amount used, and the effect on the market. Fair use is a feature of American law, not a universal one. Many countries offer only

a narrow fair-dealing or text-and-data-mining exception, and some no general defense at all, so what follows is a United States analysis. AI companies staked their defense on fair use. Taken factor by factor, it was weaker than it looked.

Factor One: Purpose and Character

The AI industry argued that training was "highly transformative."[372] Recent rulings complicated this claim.

In *Andy Warhol Foundation v. Goldsmith* (2023), the Supreme Court narrowed transformativeness. A new meaning or message is not enough on its own. What matters is whether the use serves a purpose different from the original's.[375]

The application to AI was direct. Wikipedia's purpose was answering informational queries. ChatGPT's purpose was answering informational queries. The same purpose meant weaker transformativeness claims.

The U.S. Copyright Office's May 2025 report on AI and copyright reinforced the distinction with a spectrum. Training a model for research sits at the transformative end. Training a model to generate competing outputs sits at the other.[160]

ChatGPT competed directly with Wikipedia.[21] Both answered the same questions. Both served the same users. The use was also commercial. OpenAI's revenue was running at $20 billion a year by late 2025.[286] Commercial use weighed against fair use.

Factor Two: Nature of the Work

The argument that Wikipedia was "just facts" failed against the evidence.

Wikipedia articles were not databases of raw information. They were carefully written, edited, and curated works of authorship. Editors chose what to include, how to phrase it, and which sources to cite.[82] The selection and arrangement was protected expression.

Facts themselves were not copyrightable. The expression of facts was.[376] Wikipedia's value came from that expression: the curated, verified, neutrally written summaries that made raw information accessible.

Factor Three: Amount Used

AI companies used all of Wikipedia. Not excerpts. Not samples. The entire corpus.

GPT-3's training included approximately three billion tokens from English Wikipedia. The model saw Wikipedia 3.4 times, more than any other source.[7] LLaMA used Wikipedia at 4.5% of its training data.[8]

Using the entire corpus of a work weighed against fair use. Selective quotation might qualify for protection. Total extraction did not.

Factor Four: Market Effect

Market effect carried the greatest legal weight.[377] It weighed heavily against AI companies, and the harm was documented, not argued: human traffic down 8 percent year over year once bots were filtered out,[21] organic search referrals down 26 percent over three years,[164] click-through from AI Overviews down 61 percent,[250] and zero-click searches climbing from 56 to 69 percent of all queries.[378]

Wikipedia was being substituted by AI products built on its content. Users who once visited Wikipedia now received Wikipedia-derived answers through ChatGPT and Google AI Overviews. The traffic that sustained the volunteer feedback loop was being captured by commercial systems.[21]

The substitution did not depend on verbatim copying. Even with no exact copy anywhere in the output, the model still answered the question the reader would once have brought to Wikipedia. A summary that conveys the same information is a substitute in the only sense a market recognizes. The reader who gets the answer does not make the visit, and demand for the original falls by definition, not by accident.

This is the theory lawyers call market dilution, and in May 2025 the U.S. Copyright Office endorsed it. Factor four, the report argued, should weigh whether AI systems flood a market with outputs that serve the same purpose as the works they trained on, even absent proof of harm to any single work.[160] A model trained on Wikipedia and set to answer the questions Wikipedia answers is the textbook case.

In *Kadrey v. Meta* (June 2025), Judge Chhabria warned: "It seems likely that market dilution will often cause plaintiffs to decisively win the fourth factor—and thus win the fair use question overall."[377]

Wikipedia had market harm evidence. The Foundation did not use it.

Precedents Support Enforcement

The legal terrain was not settled in AI companies' favor. By mid-2026 the first training cases had been decided, and they pointed in several directions at once. That was the point. The question was genuinely open, and open questions are what trials are for.

The "transformative" label, it turned out, did little of the deciding. In *Bartz v. Anthropic* (June 2025), Judge Alsup called training "spectacularly transformative." Anthropic settled anyway, for roughly $1.5 billion, once its use of pirated copies exposed it to damages in the hundreds of billions.[157,379,380] In *Kadrey v. Meta* (June 2025), the same adjective helped Meta win, but only because the plaintiffs brought no evidence of market harm.[377] Two courts applied one vague word to reach opposite results. What actually separated the cases was economic: whether the use displaced a market. On that question, the one the fourth factor really asks, Wikipedia had the very evidence Meta's plaintiffs lacked.[21]

Where market substitution was plain, the defense failed outright. *Thomson Reuters v. ROSS Intelligence* (February 2025), the first training case to reach judgment, held that a commercial tool built to compete with its source, using that source's own materials, was not fair use.[292] ChatGPT

was an answer engine built to compete with Wikipedia using Wikipedia's materials. The logic mapped directly.

None of this guaranteed a win, and it did not need to. *NYT v. OpenAI* showed what a rightsholder with conviction could do. Judge Stein let most of the copyright claims proceed, and by mid-2026 OpenAI was being compelled to hand over a sample of twenty million ChatGPT conversation logs in discovery.[381,382,383,384] The Foundation's claims were at least as strong, its content more central to the models, its market harm better documented. There was more than enough to survive a motion to dismiss and reach a real trial. The Foundation never filed.

CC-BY-SA Adds Claims

Even if fair use covered training, an open question, CC-BY-SA violations were systematic.

CC-BY-SA required attribution to Wikipedia, but AI companies rarely provided it in outputs. It required attribution to editors, but this was never provided. It required ShareAlike for derivatives under the same license, but model weights were proprietary. It required attribution fit to the medium, but no attribution mechanism existed at all.[16]

The Foundation had independent claims under the license.[136] Attribution was not reaching users. Citations drew clicks on one percent of visits.[385] ShareAlike was not being honored. Model weights and outputs were proprietary.

On this analysis, all four fair use factors cut against AI companies. CC-BY-SA violations were systematic. Recent rulings had begun to give content owners real footing.

The Foundation chose Enterprise revenue instead.

A Choice, Not a Verdict

Defenders of the Foundation's path made a practical case. Litigation was uncertain and expensive. AI companies were not going to stop training

on Wikipedia, sued or not. The content was freely available. The training had already happened. A lawsuit would take years, cost millions, and might fail.

Better, they argued, to engage through Enterprise and capture some value. The 30% cap on commercial revenue, which kept most funding flowing from donations, protected the Foundation's nonprofit character.[386] Partnerships with tech companies served Wikipedia's mission. If extraction was inevitable, at least collect something. The companies should, the Enterprise pitch ran, support the resource they depended on.[387]

The Foundation That Sued the NSA

The practical case has one problem: the Foundation's own docket.

On March 10, 2015, the Wikimedia Foundation filed suit against the National Security Agency.[388] It was lead plaintiff, joined by eight other organizations, from Human Rights Watch to The Nation, challenging the NSA's mass interception of internet traffic. The ACLU carried the case pro bono.[389] Jimmy Wales and Executive Director Lila Tretikov announced it in a New York Times op-ed. "Knowledge flourishes where privacy is protected," they wrote.[390] Tretikov went further: "By tapping the backbone of the internet, the NSA is straining the backbone of democracy."[388] Asked whether the Foundation was ready for a fight that could last a decade, spokeswoman Katherine Maher said it plainly: "The Foundation is prepared and committed for the duration."[391]

She meant it. The case was dismissed, appealed, revived by the Fourth Circuit in 2017 for Wikimedia alone, litigated through a state-secrets fight, appealed again, and pursued to a Supreme Court certiorari petition that was finally denied in February 2023.[392,393,394] Eight years of federal litigation against the least promising defendant in American law, on principle, with the ACLU carrying the costs.

The Foundation also picked copyright fights when it chose to. In 2014, photographer David Slater demanded that Wikimedia Commons

take down the famous "monkey selfie," the self-portrait a crested macaque had snapped with his camera. The Foundation refused, and disclosed the refusal proudly in its first transparency report: "copyright cannot vest in non-human authors," so the image had fallen into the public domain.[395,396] Slater, a working photographer, told the BBC the free copies had cost him at least £10,000 and were "killing his business."[397]

The U.S. Copyright Office later wrote the Foundation's position into its official compendium, with "a photograph taken by a monkey" as the canonical example of what no one owns.[398] The Foundation had held the line on copyright doctrine, in public, against a lone photographer, because the doctrine kept content free.

Incapacity cannot explain the reflex Chapter 12 traced through the paid-editing scandals. This was an institution that would litigate for eight years against the NSA to defend its readers' privacy, and would face down an individual photographer to defend the public domain. Privacy, openness, free content: when the cause fit its self-image, the Foundation fought, and fought well. Enforcing its volunteers' license against the companies extracting their work was a cause of a different kind. The courtroom was never the obstacle. The defendant was.

The structure explains which defendants were thinkable. The pattern matched regulatory capture. AI companies needed Wikipedia for content. The Foundation needed revenue for sustainability. Enterprise created dependency, $8.3 million and growing.[20] Dependency prevented enforcement, because you cannot sue paying customers.

Non-enforcement enabled extraction from two directions at once, the invisible kind through ChatGPT and the visible kind through Grokipedia. No villain was necessary. Institutional interest had drifted into alignment with the extractors' own, and once Enterprise revenue mattered, enforcement became structurally implausible.

And the Foundation's abstention was conspicuous precisely because so many others chose the opposite. By mid-2026, more than a hundred

copyright suits against AI developers were moving through the federal courts—the count crossed a hundred that April and stood at 114 by June—brought by a cross-section of the people whose work the models had absorbed: novelists and the Authors Guild, visual artists, news publishers, record labels, and software developers.[399] A dozen of the cases against OpenAI had been consolidated into a single multidistrict proceeding in April 2025.[400] Many of these plaintiffs had weaker claims than the Foundation, and thinner evidence of market harm. They sued anyway.

The sharpest single comparison was the New York Times. The Times faced the same uncertainties, the same training-is-transformative arguments, the same industry pressure. Wikipedia's content was arguably more valuable to the models, the factual backbone of ChatGPT's knowledge. The Times sued OpenAI in December 2023 and pressed toward trial.[381]

The Foundation answered the same era with Enterprise contracts and AI-company revenue.[19] One chose enforcement. One chose accommodation. The difference was not the legal terrain.

The moments to act were specific. In May 2020, GPT-3 disclosed its Wikipedia training.[7] In November 2022, ChatGPT launched. In December 2023, the Times sued OpenAI.[381] In March 2024, 273 Ventures, a legal-AI startup, asked the Foundation how to comply with the license.[18] In February 2025, Thomson Reuters won the first AI training case.[292]

At each point the Foundation could have demanded attribution, required licensing, joined or paralleled the litigation, or published an enforcement position. It did none of these things.

Why No Case Was Made

Copyleft did not fail because the license was inadequate. It failed because the institution's immune response was disabled from within.

The mechanics were the ones Chapter 16 laid out: a staff that gave almost entirely to one party, a leadership drawn from the same professional

world as the executives across the table, and a revenue line that turned would-be defendants into partners.[22] Culture and money pulled in the same direction, and no one inside was placed to pull against them.

Stallman's license worked as designed. Enforcement required adversarial instincts that the monoculture had bred out of the institution.

The Unkept Promise

If enforcement was possible, and the evidence says it was, then non-enforcement was a choice. Not passive drift. Not institutional inertia. A decision made at specific moments by specific people who judged that Enterprise revenue mattered more than licensing integrity.

That decision had a cost, and the people who paid it never agreed to it. Volunteers built Wikipedia under the promise of CC-BY-SA, and a quarter million of them still maintain it.[365] Their work would stay free. Derivative works would be shared alike. Attribution would follow their words wherever those words traveled.[16]

No one asked those volunteers when the Foundation signed Enterprise deals. No one asked them when the Foundation admitted, in 2024, that most of its customers were not complying with the license.[18] No one represented them when revenue won out over the license they had been promised.

The case for enforcement was there to be made. The people who could have made it answered to customers now, not to the contributors whose promise went unkept.

CHAPTER TWENTY-TWO

The Volunteer Question

> Individually, there's only so much that I can do.
>
> — Robert Sim, 2025 Wikimedian of the Year

For two decades, Robert Sim documented Singapore on the English Wikipedia. Its culture, its history, its politics: he wrote and corrected and maintained it, edit after edit, past 79,000 of them.[401] He had watched the project grow from curiosity to institution. He had stayed when others left. The Wikimedian of the Year award that marked Wikipedia's twenty-fifth year was recognition of that endurance.

His response to the honor was modest. Individually, there was only so much he could do, he said. He wanted to involve more people.[401]

The hope was genuine. The condition behind it was demanding. Wikipedia runs on people. As the Foundation marked the anniversary with record revenue from Wikimedia Enterprise, its paid service selling Wikipedia's content to AI companies, one question hung over the celebration.[20,315] Why would anyone volunteer their labor when that labor trained competitors?

The Feedback Loop

For twenty-five years, Wikipedia sustained itself through a feedback loop. That loop was a commons-making mechanism.

A reader visits Wikipedia. They notice an error: a misspelling, a missing fact, an outdated figure. They click edit. They fix the problem. They watch their correction go live, verified, read by millions. They feel the satisfaction of having added to something owned in common.

That moment makes a commons. Not the infrastructure, not the license, not the steward. The commons is the quarter million people who volunteer because they believe they are building something that belongs to everyone equally.[13]

When that belief breaks, the commons dies.

The loop converted readers into editors. The simplicity of "anyone can edit" was both ideology and recruitment. Wikipedia grew from a handful of contributors in January 2001 to active editors in the hundreds of thousands across its languages by its twenty-fifth anniversary.[47,13]

The motivations were well documented. Research identified intrinsic factors such as enjoyment and learning, social factors such as community and identity, and values such as ideology and public service.[402] Recognition mattered measurably. A 2012 study found that informal awards sharply increased editors' later productivity.[403] Volunteers continued when they felt part of something bigger than themselves.

The loop worked because the feedback was immediate and tangible.

Henner's Warning

Christophe Henner, former chair of the Foundation's Board, published a "wake-up call" on the twenty-fifth anniversary.[404] His assessment was blunt:

> *The feedback loop that has sustained volunteer motivation for 25 years is simple: I write, people read, I can see the impact. Break that loop and you break the engine for some contributors.*

The question was whether anything could repair the loop.

The Long Decline

The volunteer crisis began before AI. Editor counts on the English Wikipedia peaked in 2007 at roughly 51,000 monthly active editors, those making five or more edits. By 2014, the count had fallen to between 30,000 and 35,000, where it stabilized.[192]

Research identified the causes. Bots automated the gatekeeping.[192] A ruleset that began as a handful of core principles grew to hundreds of policies and guidelines, plus thousands of essays frequently invoked in disputes.[91,405] A hostile culture "bit" newcomers.[406] A mobile gap made editing from phones difficult.[407]

Aaron Halfaker's 2013 study documented the paradox.[192] The very systems built to maintain quality were now preventing growth. The filters that kept vandalism out also kept new editors out.

The decline stabilized but never reversed.[192] Wikipedia became, in Henner's phrase, "a smaller club working harder."[404]

AI Acceleration

Then came ChatGPT.

In April 2025, ChatGPT overtook Wikipedia in monthly visits in the United States.[408] The model built from Wikipedia's text now stood between Wikipedia and its readers. The irony was precise. Wikipedia supplied nearly half of ChatGPT's cited sources, and ChatGPT was replacing it.[12]

The traffic data confirmed the shift:

Metric	Value
Human traffic decline (after bot detection)	-8% YoY[21]
Organic search traffic decline (3-year)	-26%[164]
Zero-click searches (2024 to 2025)	56% to 69%[409]
New registrations decline (2016-2025)	-36%[404]

Table 22.1. Wikipedia traffic decline metrics

The feedback loop was breaking at both ends. Fewer readers meant fewer potential editors. A rising share of searches were now zero-click, answered on the results page so users never visited the source, and AI answers meant they never reached Wikipedia at all.[21]

The Broken Contract

The crisis was not only about declining numbers. It was about a promise broken.

Wikipedia volunteers contributed under an implicit bargain. Their labor would remain free and accessible. It would be governed by the community. The copyleft license would ensure that anyone building on their work shared improvements back.[16] They were building a commons.

In 2024 and 2025, that bargain visibly shattered. Volunteers saw their words reproduced in locked proprietary systems. They saw users asking ChatGPT instead of visiting Wikipedia.[164] When the Foundation signed its Enterprise data-feed deals,[29] volunteers discovered they had not been consulted. The institution they had built had grown dependent on the companies extracting their labor.

When the Foundation tested AI summaries in June 2025, volunteers killed the experiment within a day, and they rejected Wales's proposal to use ChatGPT for article reviews the month after.[24] They were not opposed to AI in principle. They were opposed to being extracted without consent.

The volunteer question was no longer "why are people leaving?" It was "why would anyone stay?" The motivations that sustained contribution were being hollowed out. AI removed the sense of impact, because volunteers could not see their work when a chatbot answered. Enterprise undermined community, because the Foundation took money without consultation. The harvesting betrayed the promise of building something lasting, because the labor now trained its own replacement.

The Voices

Ian Ramjohn at Wiki Education put the threat in a sentence: "Definitely LLMs are an existential threat to Wikipedia."[410]

Scientific American traced the mechanism in its anniversary coverage:

> *Fewer visitors mean fewer new editors for Wikipedia, and less frequent visits mean slower correction of errors… If fewer people visit and fewer people edit, the system that made Wikipedia self-correcting—and unusually resilient—could weaken.*[410]

Wagner and Jiang, writing in the *Journal of the Association for Information Science and Technology*, modeled the endpoint as a "vicious cycle":

> *Contributors who reduce their contribution efforts as AI pervades the platform will leave Wikipedia increasingly dependent on additional AI activity… [causing] a vicious cycle leading to a staling of the content and diminishing value of this venerable knowledge resource.*[411]

The logic was a closing loop. Volunteers withdraw. Wikipedia grows more AI-dependent. Quality declines. The resource degrades.

Against that reading, the community mounted a defense. WikiProject AI Cleanup, formed in December 2023, had tagged nearly 3,000 articles for AI contamination by October 2025.[343,342] Its members built detection methods, documented patterns, and won concrete rules, including one that let editors delete obvious AI-generated text on sight.[344,342] The same community had killed the summaries experiment and rejected Wales's ChatGPT proposal.[24]

The editor Bawolff had put the objection crisply: Wikipedia's brand was reliability, traceability, and the promise that anyone can fix it, and AI was the opposite of all three.[24]

Ilyas Lebleu of WikiProject AI Cleanup made the stakes concrete: "When readers come to Wikipedia, they expect something that was writ-

ten by volunteers, that was checked by volunteers. They don't expect something that was written by an AI."[342]

The defense was principled. Whether it was sufficient stayed uncertain.

And some volunteers simply stayed. Sim was one of them: two decades of work and a commitment to free knowledge that the harvesting had not dissolved.[401] The community bonds held. The work still mattered. These were the editors maintaining Wikipedia at twenty-five, the dedicated core, moved by values that commercial extraction could not entirely undermine. They continued because the alternative, abandoning the collective work, was worse. Whether values alone could sustain contribution, once every other motivation had been eroded, was the question none of them could answer.

Sim's Wish

Sim's wish, to involve more people, carried its full weight.[401] For Wikipedia to continue, it needed exactly what he had asked for. The same conditions that made him ask gave others their reasons to leave.

The arithmetic was plain. Enterprise revenue reached $8.3 million in fiscal 2024-25.[20] The volunteers who wrote the content that made that revenue possible received none of it. The people who create the resource that powers AI receive nothing from the commercial relationships their labor enables.

The volunteer question was not really about Wikipedia. It was about collective work itself. What happens when the people who *are* the shared resource realize it is being harvested?

The issue is governance. If the people who create a thing cannot control what happens to it, it is not a public good. It is exploitation with a free license.

The volunteers who maintain Wikipedia have no answer to this question. Neither does the Foundation. Neither does anyone.

For twenty-five years, volunteers contributed because they believed they were building something that would remain free, accessible, and theirs. The AI era revealed how much that belief depended on a Foundation willing to defend it. Their labor was valuable enough for billion-dollar companies to train on and for the Foundation to monetize.[315] It was not valuable enough to compensate, or even to consult.

Wikipedia needs people. People need reasons. The feedback loop that sustained the encyclopedia for a quarter century is breaking, and what remains when the builders have every reason to walk away is the question that will not resolve on its own.

CHAPTER TWENTY-THREE

Four Futures

> Neither the state nor the market is uniformly successful in enabling individuals to sustain long-term, productive use of natural resource systems.
>
> — Elinor Ostrom, *Governing the Commons*

In January 2026, Wikipedia looked exactly as it always had. Anyone could open an article and change it, free of charge.[6] The economics underneath had inverted. Six of the largest AI companies now paid the Foundation for the content volunteers still gave away.[19] Human traffic had fallen 8 percent in a year.[21] New registrations were down 36 percent from their 2016 peak.[404]

These were the documented starting conditions, and the community that read them had already split four ways over what they meant. The pragmatists saw inevitability: AI was here, so the Foundation should be paid for access. The copyleft purists saw the founding promise being broken. The skeptics saw an existential threat in AI itself. The volunteer advocates saw their labor turned to someone else's profit without consent. Each camp read the same numbers and saw a different path forward. None commanded consensus.

Factor	Current State
Enterprise revenue	$8.3 million (FY 2024-25)
AI customers	Amazon, Meta, Microsoft, Google, Perplexity, Mistral[19]
Human traffic	Down 8% YoY
New registrations	Down 36% since 2016
Community factions	Four-way split
Legal status	NYT v. OpenAI in discovery; Thomson Reuters v. ROSS on appeal[381,412]

Table 23.1. Wikipedia status as of January 2026

The question was not whether Wikipedia would survive. The encyclopedia would persist in some form. Whether a commons could survive capture was the harder question.

What follows are four futures. They are not predictions but pictures, each built by following one of the forces already in motion to the end of its logic. The real future will braid strands of all four. Setting them side by side is a way to see what each would cost, and to ask the questions the confident version of each story likes to skip.

Scenario One: Wikipedia Captured

The Enterprise path continues. Revenue grows from $8.3 million to $15 million to $50 million. Every major AI company pays for access. The Foundation reinvests. Individual access remains free.

Wikipedia becomes infrastructure for the AI industry, a public good supported by those who profit from it. Volunteers maintain content while commercial entities extract value. The copyleft ideal becomes a dead letter.

The Foundation achieves financial sustainability. AI companies gain legitimate access. Users receive AI answers based on verified content, at first. But volunteers see their labor absorbed without compensation or

consultation. The ShareAlike promise dies. Future open projects learn that free content means free for harvesting.

Once revenue reaches $50 million, the Foundation cannot risk losing it. When AI companies request changes to terms, data formats, or attribution, the Foundation must weigh community values against commercial relationships. The entity that should advocate for the shared resource becomes the entity that negotiates its harvesting.

The word doing the reassuring here is "public." It deserves suspicion. A public good funded by the companies it serves is not obviously public, and the payers are not a cross-section of the public but a handful of the most powerful firms alive. As their money becomes indispensable, the uncomfortable questions arrive quietly. Would an unflattering edit to the article on OpenAI or Google be handled the way an unflattering edit to any other corporation is? The Foundation would insist its content stayed walled off from its contracts. It would also have every reason never to test the claim.

What survives is a database. What dies is the commons.

Scenario Two: Wikipedia Fractures

A critical mass of editors concludes the Foundation has betrayed copyleft. They announce a fork, invoking CC-BY-SA's permission to copy and modify.[16]

The fork implements strict attribution requirements. Technical measures block non-compliant scrapers. Some volunteers migrate. Most stay with the original. Two communities develop, each claiming to represent Wikipedia's true values.

There was a precedent. Oracle acquired Sun Microsystems in 2010, becoming the steward of OpenOffice.org, the free office suite that rivaled Microsoft Office. Developers who doubted Oracle's commitment forked the project into LibreOffice. Research found the fork attracted "the long-

term and most active committers" and achieved a "long-term sustainable community."[413]

But Wikipedia is not software. Articles require maintenance across millions of pages.[6] The volunteer base cannot sustain two encyclopedias, and a fork would divide an already-declining contributor pool.[192] Wikipedia's own history offered the warning: the 2002 Spanish fork, Enciclopedia Libre, never matched the original's scale[57] and went dark in 2024.[62]

Editors committed to strict copyleft would preserve their principles in at least one version. But both encyclopedias would face divided resources and duplicated effort. Readers would face confusion about which Wikipedia to trust. The movement's civil war would consume energy that could build knowledge.

And a fork would not touch the free-riders it was meant to punish. The companies that never signed anything, and simply scraped the public pages, would scrape the fork just as freely. Strict copyleft on a second site changes the rules only for those already willing to follow them.

What survives is two partial projects. What dies is the unified commons.

Scenario Three: Wikipedia Enforces

Courts rule against AI fair use. The *NYT v. OpenAI* trial proceeds in late 2026 and finds that training without license is infringement. The Thomson Reuters precedent is affirmed on appeal.

AI companies face a choice: license content or remove it from training. A publishers' coalition demands compliance. The Wikimedia Foundation, pressured by community and courts, joins the enforcement effort.

The legal foundation exists. The Copyright Office's May 2025 report identified market effect as the single most important element of fair use and warned of significant potential harm from AI training.[160] Judge Vince Chhabria, ruling in *Kadrey v. Meta*, warned that market dilution could lead plaintiffs to win the fair use question in many AI training cases.[377]

Content creators gain licensing revenue and attribution. The copyleft ideal is enforced against AI. Volunteers see their labor finally acknowledged. AI companies face billions in licensing costs, and users receive degraded or more expensive AI outputs.

Enforcement would carry its own complications. The Foundation admitted in March 2024 that it suspected customers of non-compliance, yet the Enterprise deals continued.[18] Courts may find it implicitly ratified AI training through years of accommodation. Enforcement now would require explaining that silence. Revenue relationships create conflicts of interest.

What survives is the commons, and the institution that abandoned it. What dies is the Foundation's claim to be Wikipedia's true steward.

Scenario Four: Wikipedia Declines

No dramatic resolution. The status quo continues until it doesn't.

Registrations keep declining. Traffic keeps falling. ChatGPT keeps drawing the queries that once went to Wikipedia. The smaller club working harder exhausts itself. Article quality degrades imperceptibly, then visibly.

AI systems trained on 2024 Wikipedia increasingly cite outdated or degraded content. The feedback loop closes: degraded AI cites degraded Wikipedia cites degraded AI.

The contamination is already visible. ChatGPT has begun citing Grokipedia,[28] the AI-built encyclopedia that copied Wikipedia and laced it with extremist sources.[26,32]

The brand itself has eroded. "Is 'I edit Wikipedia' something people say with pride or embarrassment?" former Board chair Christophe Henner asked. "Contributing to open source on GitHub has cachet. Making TikToks has cachet. Editing Wikipedia? We've become the encyclopedia your teacher warned you about, not the movement you want to join."[404]

No one profits from this version for long. This is the tragedy of the commons realized. Everyone loses. Readers lose reliable information. AI loses quality training data. The knowledge commons degrades for all.

This decline is not fated, and it has been forecast before. Researchers pronounced Wikipedia's editor base in terminal decline around 2011, and the encyclopedia kept going for another fifteen years.[192] A smaller, more committed corps might yet stabilize it, and a web drowning in AI slop might even send readers back toward a source that still shows its work. Nothing guarantees the worst case. It requires only that no one act to prevent it.

What survives is a degraded database. What dies is the community, and its belief that the work was ever collectively owned.

The law had not run out. The legal case for enforcement existed: the copying was documented, the market harm measured, the precedents developing. What ran out was the will of the institution entrusted to use it.

Resources were not the constraint. The Foundation held hundreds of millions in revenue and reserves.[23] What it lacked was appetite for the fight, because the would-be defendants had become its partners.

The scenarios are not alternatives but simultaneous trajectories. Their relative strength will decide what comes. By mid-2026 the third had its first real test on the docket: the Third Circuit heard the first federal appellate argument on fair use in AI training that June, in the ROSS appeal, with the decision pending.[414]

The futures were still arguments when the institution revealed its character. On June 22, 2026, as Chapter 3 recounts, the community blocked Larry Sanger, the co-founder who had named the encyclopedia.[30,36] The block had a stated basis. But the same institution that could not summon the will to enforce its license against companies worth trillions summoned it, within days, against a critic whose only power was his voice. The

institution keeps its conviction for the weak and its accommodation for the strong. Whichever scenario prevails, that character will shape it.

Elinor Ostrom won the Nobel Prize for showing that communities can govern shared resources without the state or the market.[415] Her eight design principles explained why some commons endured and others collapsed: clear boundaries, rules fit to local conditions, collective decision-making, monitoring, graduated sanctions, conflict resolution, the recognized right to self-organize, and nested layers of governance.[416] Wikipedia met nearly all of them.

It met them for content. The governance that curated articles never reached the commercial relationships. Enterprise contracts were confidential. Volunteers were not consulted.

Ostrom's principles assume the institution itself stays uncorrupted. They offer no defense when the guardian becomes the extractor. That is the underlying failure, and it is not a failure of commons governance. It is the limit of commons governance against capture from within.

The capture did not require bad actors. It required only that institutional interests align with the harvesters', and they did. Enterprise launched, revenue grew, the AI companies became partners.[38,20,19] The Foundation acknowledged suspected non-compliance and kept selling.[18] Once the revenue mattered, legal action became inconvenient, then unthinkable.

The pattern is not unique to Wikipedia. Open source software has drifted from copyleft to permissive licenses.[417] Copyleft's share of projects fell from a majority in the early 2010s to a minority a decade later.[418,417] Creative Commons, which built ShareAlike, took the position in 2021 that AI training should be considered non-infringing by default, and its 2025 guidance qualified that stance without reversing it.[69] Create a shared resource, license it freely, and watch the institutions meant to protect it grow dependent on those who harvest it.

Copyleft was built to keep knowledge free through law.[107] The protection it promised did not hold.

It failed because legal interpretation is contested. Courts may yet rule that AI training is fair use, or that statistical weights are not derivative works. The mechanism Stallman designed cannot function if judges read it away.[2]

It failed because technology raised the cost of proof. Training folds text into model parameters, where copying must be demonstrated rather than seen. The copies remain in the weights, as the memorization studies showed. Extracting them takes an enforcer willing to look, and none appeared.

It failed because a license without enforcement is only a wish. The entity that should have enforced it chose revenue instead.

And it failed because the Foundation had no internal opposition. Of the political money identifiable in federal election records, more than 99 percent went to a single party—just $151, across eighteen years, went to Republicans.[22] The party that received the money was also the party of the companies whose compliance the Foundation was supposed to be demanding. The people inside shared the harvesters' networks and assumptions, so no faction argued for confrontation. Partnership felt natural to people who saw tech executives as colleagues.

The book began with a different problem. Nupedia's experts could not produce free knowledge fast enough, and Wikipedia answered with radical openness: anyone can edit.[36] Twenty-five years on, the problem has inverted. The question is no longer how to make knowledge free. It is how to keep it free. The openness that allowed creation also allowed ingestion.

Forty years after Symbolics taught Stallman what unshared improvements cost,[2] the AI companies ran the same play on Wikipedia: they took the volunteers' work, trained their models on it, and locked the results

in proprietary systems. ShareAlike was ignored. Attribution was never provided. The Foundation knew. It signed the Enterprise deals anyway.[18]

The license did not fail. The institution failed.

Nothing here is redemption, because none has occurred. The outcome still depends on choices not yet made, by courts, by the Foundation, by the people who keep editing. The encyclopedia remains. The license remains. The Foundation remains. But "free" has become something else: free for harvesting, free for training, free for the commercial systems that compete with the resource they consume. That is the story of Wikipedia, and the lesson of the commons.

Somewhere tonight, an editor with tens of thousands of edits, twenty years of contribution, and articles that would not exist without their work is weighing whether to open the editing window again. Their labor has trained ChatGPT, Google's AI Overviews, and a dozen other systems. None of them says so.

Whether that editor still feels part of a shared project, or feels their work has simply been absorbed, will decide more than any court ruling.

On a quarter million screens, the weighing is under way.

Epilogue

In the early 1980s, Richard Stallman stood before a dead machine and wept. The PDP-10 had been more than hardware. It was the center of a community of programmers who shared code freely, and when Symbolics took that shared work and refused to give it back, the community scattered and the machine went dark.[2]

From that grief came copyleft: use my work, but share your derivatives. No one could ever again take what the community built and lock it away.[2]

In 1999 Stallman proposed extending the idea to knowledge itself, and two years later Wikipedia launched under a copyleft license.[1,47] Twenty-five years on, the encyclopedia had outgrown his vision, seven million English articles, a quarter-million editors, billions of readers, and the protection underneath it had quietly failed.[13]

Not because the license was weak. Because the institution entrusted with it chose not to use it.

The evidence had been in the Foundation's own hands since March 2024, when it admitted in writing that many AI companies "may not be compliant with the letter of the Creative Commons rules or the spirit of the licenses."[18] It kept signing deals regardless.[29] When Grokipedia appeared, a visible fork that at least carried attribution and free access alongside its neo-

Nazi citations, the Foundation condemned the content while partnering with companies whose quieter violations ran deeper.[25,26,27]

It held the resources to fight: more than $180 million a year and hundreds of millions more in reserve.[419,185] What it lacked was the will, and the independence, to turn on companies that had become its partners. That independence had gone first: a political monoculture captured the Foundation from within before the AI industry captured the commons from without. The license did not fail. Its guardian did.

Yet this is not where the story ends.

The volunteers remain. About 250,000 editors still maintain Wikipedia every month, still fixing errors, adding citations, arguing on talk pages about neutrality and sourcing.[13] They still believe their contributions matter, and their choice will decide the outcome.

If they conclude the promise was broken and drift away, the database will persist but the community will scatter. Another machine will go dark. If instead they organize, protest, fork, or simply refuse to accept that extraction is inevitable, the Foundation will have to answer. The Spanish editors who forked in 2002 forced Jimmy Wales to abandon his advertising plans.[57] Community pressure has bent Wikipedia's course from the start. The volunteers are not powerless. They have not yet chosen to use their power.

Stallman's answer to Symbolics was copyleft: keep shared work free, forever. In 2026 a different question presses. When the institution entrusted to enforce that mechanism is captured, what then?

The copyleft promise was never self-enforcing. It required humans willing to defend it. Stallman fought Symbolics for two years, alone,

because he believed the community was worth saving. The MIT hackers did not fight with him. They drifted to industry jobs, and the community died.

Wikipedia's volunteers face the same choice. They can drift away, accepting that their labor will train the systems that replace them. Or they can fight, demanding enforcement, transparency, accountability.

The commons does not die when the license fails. It dies when the people who built it decide their contributions no longer matter.

Stallman's grief was never about a machine. It was about a community that would not fight to save itself.

The volunteers who edit Wikipedia today are not yet at that moment. The community still exists. The work still matters. The question of whether to fight remains open.

No outsider can answer it. Only the volunteers can.

But we can say what is at stake: not just an encyclopedia, but the belief that collective ownership of knowledge is possible. That belief survives only if someone defends it.

The commons was never the database. It was the decision, made millions of times by hundreds of thousands of people, that contributing to something shared was worth the effort.

That decision is made fresh every day.

A Personal Note

A reader deserves to know where a writer stands. I did not come to this story as a neutral observer, and it would be dishonest to pretend otherwise.

I have been part of free-software and open-knowledge communities since 1999, the same year this book's story begins, when Richard Stallman first imagined an encyclopedia that could not be enclosed. The principle at the center of these pages is one I have believed for most of my adult life: that knowledge built in common should stay in common. It is also the work I do now. Through the ALEA Institute, a non-profit I help lead, I work on open research and education around the ethical development of AI.

That work is why I appear, quietly, in the record. The March 2024 exchange that recurs throughout this book, in which the Foundation admitted that many AI companies likely did not honor its license, began with a question my colleagues and I asked. We wrote to the Wikimedia Foundation to learn how to train a model in compliance with CC-BY-SA. Chapter 16 tells that story in the third person. The company it names was mine. I sent those emails, and my wife, Jillian Bommarito, published the account this book cites.

I kept the account in the third person not to hide it but because the argument should rest on the documents rather than on my telling. The emails say what they say. A reader can weigh them without trusting me.

Still, the connection should be stated plainly. I have never edited Wikipedia or worked for the Foundation, and my only direct part in the events of this book runs through that exchange. But being the one who asked gave me an early and close view of how the Foundation treated the only company that asked, and a reader is entitled to wonder whether that has colored my judgment. I have tried not to let it. I have cited every claim I could and left the record open, so that you can check me where I fall short.

I want the commons to survive. That hope is the reason for this book, and it is the bias worth weighing while you read.

Michael J Bommarito II

A Note on Sources and Methods

Writing the history of Wikipedia presents unique challenges and opportunities. The project's radical transparency means that nearly every decision, dispute, and discussion has been preserved in searchable archives. This abundance of primary sources offers unusually broad access to how the encyclopedia actually developed—but also requires careful sorting.

Source Hierarchy

I have organized sources into tiers based on reliability and proximity to events:

Tier 1: Wikimedia Foundation Documents. Board resolutions, annual reports, Form 990 tax filings (the annual IRS disclosures filed by nonprofits), and official statements provide authoritative information about the Foundation's operations, finances, and governance decisions.

Tier 2: Wikipedia Primary Sources. Policy pages, talk page discussions, arbitration cases, and edit histories document how the community actually functioned. These sources are invaluable but require careful interpretation—they show what people said publicly, not necessarily what they thought privately.

Tier 3: Academic Research. Peer-reviewed studies of Wikipedia provide systematic analysis of editing patterns, demographic composition, and content quality. I have relied heavily on work by researchers at the

Oxford Internet Institute, MIT, and the Wikimedia Foundation's own research team.

Tier 4: News Coverage. Contemporaneous journalism captures how events were understood at the time, though later reporting sometimes repeats earlier errors.

Tier 5: Memoirs and Interviews. First-person accounts provide valuable perspective but often reflect later rationalizations. I have used these cautiously, preferring documented evidence over remembered claims.

Verification Practices

For claims about specific events, I have sought multiple independent sources. For claims about living persons, I have applied heightened scrutiny and, where possible, sought their response.

All web sources have been archived using the Wayback Machine. Wikipedia citations include revision IDs to ensure verifiability even as articles change.

The FEC Donation Analysis

One recurring claim rests on an analysis of Federal Election Commission records. Itemized individual contributions by self-identified Wikimedia Foundation employees between 2008 and 2026 total roughly $295,000 from about 240 donors. Of that sum, 99.3 percent went to Democratic candidates and committees, and roughly $1,860 to third-party or independent candidates; just $151, from three donors across the eighteen-year span, went to Republican candidates and committees.[22] The analysis matches FEC individual-contribution records on the self-reported employer field and donor name, with contributions routed through conduits such as ActBlue and WinRed attributed to the conduit's party.

Its limitations should be understood. FEC records capture only itemized contributions, generally those exceeding $200 to a single committee in a cycle. Employer self-reporting is inconsistent, and name matching can

produce both false positives and false negatives. The pattern is striking, but it stands as evidence about identifiable donations, not about the private politics of every employee.

Limitations

The account has three main limits:

This history focuses primarily on the English Wikipedia and the Wikimedia Foundation. The experiences of other language communities, while occasionally referenced, deserve their own treatment.

Internal Foundation communications and private discussions among editors remain largely inaccessible. The picture presented here is necessarily incomplete.

Wikipedia continues to evolve. Events described here should be understood in their historical context, not as descriptions of current practices.

My own proximity to these events is disclosed in "A Personal Note," which precedes this appendix.

Bibliography

[1] Richard M. Stallman. *The Free Universal Encyclopedia and Learning Resource*. Free Software Foundation. 1999.

[2] Sam Williams. *Free as in Freedom: Richard Stallman's Crusade for Free Software*. O'Reilly Media, Mar. 2002.

[3] Richard M. Stallman. "The GNU Manifesto". In: *Dr. Dobb's Journal of Software Tools* 10.3 (Mar. 1985), pp. 30–35.

[4] Andrew Lih. *The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia*. New York: Hyperion, 2009.

[5] *Why Publishers should Use the GNU FDL - GNU Project - Free Software Foundation*. 2026.

[6] Wikipedia contributors. *Wikipedia:About*. 2026.

[7] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020.

[8] Hugo Touvron et al. "LLaMA: Open and Efficient Foundation Language Models". In: *arXiv preprint arXiv:2302.13971* (2023).

[9] Nicholas Carlini et al. "Extracting Training Data from Large Language Models". In: *30th USENIX Security Symposium*. 2021, pp. 2633–2650.

[10] Milad Nasr et al. "Scalable Extraction of Training Data from (Production) Language Models". In: *arXiv preprint arXiv:2311.17035* (2023).

[11] A. Feder Cooper et al. *Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models*. 2025.

[12] *AI Platform Citation Patterns: How ChatGPT, Google AI Overviews, and Perplexity Source Information*. 2026.

[13] Wikipedia contributors. *Wikipedia celebrates 25 years of knowledge at its best*. 2026.

[14] Bloomberg Intelligence. *Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds*. 2023.

[15] *Wikimedia Licensing Policy Change – A Conundrum*. 2026.

[16] *Legal Code - Attribution-ShareAlike 4.0 International - Creative Commons*. 2026.

[17] Wikimedia Foundation. *Wikilegal/Copyright Analysis of ChatGPT - Meta-Wiki*. 2026.

[18] Jillian Bommarito. *Can You Train an LLM on Wikipedia? Wikimedia's Surprising Opinion — Jillian Bommarito*. 2026.

[19] Wikipedia contributors. *New Wikimedia Enterprise Partners: Wikipedia's 25th Birthday*. 2026.

[20] Wikimedia Foundation. *Wikimedia Enterprise Financial Report: Fiscal Year 2024 – 2025 – Diff*. 2026.

[21] Wikipedia contributors. *Wikipedia says traffic is falling due to AI search summaries and social video | TechCrunch*. 2026.

[22] Federal Election Commission. *Itemized Individual Contributions (Schedule A), OpenFEC API*. 2026.

[23] Wikipedia contributors. *Wikipedia:Fundraising statistics*. 2026.

[24] Wikipedia contributors. *Jimmy Wales Says Wikipedia Could Use AI. Editors Call It the 'Antithesis of Wikipedia'*. 2026.

[25] Plagiarism Today. *Grokipedia: How Not to Make an Encyclopedia*. 2025.

[26] NBC News. *Elon Musk's Grokipedia cites neo-Nazi website 42 times: study*. Nov. 2025.

[27] Agence France-Presse. *Musk's Grokipedia leans on 'questionable' sources, study says*. Nov. 2025.

[28] The Guardian. *Latest ChatGPT model uses Elon Musk's Grokipedia as source, tests reveal*. 2026.

[29] *Wikimedia Foundation announces new AI partnerships with Amazon, Meta, Microsoft, Perplexity, and others | TechCrunch*. 2026.

[30] Jason Koebler. *Wikipedia Cofounder Larry Sanger Banned From Site for 'Canvassing'*. June 2026.

[31] PBS NewsHour. *Elon Musk launched Grokipedia. Here's how it compares to Wikipedia*. Oct. 2025.

[32] Harold Triedman and Alexios Mantzarlis. *What did Elon change? A comprehensive analysis of Grokipedia*. 2025.

[33] Wikipedia contributors. *Wikipedia Is Badly Biased – LarrySanger.org*. 2026.

[34] Wikipedia contributors. *Grokipedia*. 2026.

[35] Grokipedia. *American History X*. 2026.

[36] Larry Sanger. "The Early History of Nupedia and Wikipedia: A Memoir". In: *Open Sources 2.0: The Continuing Evolution*. Ed. by Chris DiBona, Mark Stone, and Danese Cooper. O'Reilly Media, Nov. 2005, pp. 307–338.

[37] Wikimedia Foundation. *Wikimedia Foundation staff and contractors*. 2026.

[38] *Wikimedia Enterprise – Open Future*. 2026.

[39] Wikipedia contributors. *Wikipedia:Administrators' noticeboard/Incidents/Possible off-wiki canvassing by User:Larry Sanger*. 2026.

[40] Larry Sanger. *Well, that's that—I've been blocked by Wikipedia 'indefinitely'*. June 2026.

[41] Zachary Schwartz. *Wikipedia's Co-Founder Is Wikipedia's Most Outspoken Critic*. Vice. Nov. 2015.

[42] Robin Miller and Jimmy Wales. *Wikimedia Founder Jimmy Wales Responds*. July 2004.

[43] Joseph Janes. *Nupedia (Wikipedia predecessor)*. University of Washington Information School. 2017.

[44] Wikipedia contributors. *Bomis*. 2026.

[45] Jr. Reagle Joseph M. *Good Faith Collaboration: The Culture of Wikipedia*. History and Foundations of Information Science. Cambridge, MA: The MIT Press, 2010.

[46] Richard M. Stallman. *The Free Universal Encyclopedia and Learning Resource*. GNU Project, Free Software Foundation. 1999.

[47] Wikipedia contributors. *History of Wikipedia*. 2026.

[48] F. A. Hayek. "The Use of Knowledge in Society". In: *The American Economic Review* 35.4 (Sept. 1945), pp. 519–530.

[49] Thomas J. DiLorenzo, ed. *Hayek for the 21st Century: Essays in Political Economy*. Auburn, Alabama: Mises Institute, 2025.

[50] Eric S. Raymond. "The Cathedral and the Bazaar". In: *First Monday* 3.3 (Mar. 1998).

[51] Wikipedia contributors. *Nupedia*. 2026.

[52] *Wiki Wiki Web*. 2026.

[53] CNN. "Wikipedia's first ever edit sells as NFT for $750K". In: *CNN Style* (Dec. 2021).

[54] Wikipedia contributors. *UseModWiki*. 2026.

[55] Larry Sanger. *Why I Am No Longer Participating in the Wikipedia Project*. 2002.

[56] Wikipedia contributors. *Atonality (article revision history)*. 2002.

[57] Edgar Enyedy and Nathaniel Tkacz. "'Good Luck with Your WikiPAIDia': Reflections on the 2002 Fork of the Spanish Wikipedia". In: *Critical Point of View: A Wikipedia Reader*. Ed. by Geert Lovink and Nathaniel Tkacz. INC Reader 7. Amsterdam: Institute of Network Cultures, 2011.

[58] Wikimedia Foundation. *Spanish Fork - Meta-Wiki*. 2026.

[59] Wikipedia contributors. *Wikipedia en español*. 2026.

[60] Larry Sanger. *Jimmy Wales on Advertisement*. Jan. 2011.

[61] Jimmy Wales. *Re: [Wikipedia-l] New to the Wikipedia with a few questions*. Feb. 2002.

[62] Wikipedia contributors. *Enciclopedia Libre Universal en Español*. 2026.

[63] Jimmy Wales. *[Wikipedia-l] Advertisements*. Mar. 2002.

[64] ClickZ News. *No Ads in Wikipedia Says Wales*. 2006.

[65] Academic Dictionaries and Encyclopedias. *Enciclopedia Libre Universal en Español*. 2026.

[66] Wikipedia contributors. *Wikimedia Foundation*. 2026.

[67] Wikipedia contributors. *Wikipedia:Wikipedia Signpost/2005-04-18/Tax exempt status*. 2026.

[68] Wikipedia contributors. *Spanish Wikipedia*. 2026.

[69] *The Impact of Share Alike/CopyLeft Licensing on Generative AI – Open Future*. 2026.

[70] Larry Sanger. *My Resignation–Larry Sanger*. Mar. 2002.

[71] Larry Sanger. *My Role in Wikipedia (links)*. 2019.

[72] Peter Meyers. "Fact-Driven? Collegial? This Site Wants You". In: *The New York Times* (Sept. 2001).

[73] Wikipedia contributors. *Edit history: Jimmy Wales*. 2026.

[74] Larry Sanger. *Why Wikipedia Must Jettison Its Anti-Elitism*. Dec. 2004.

[75] Judith S. Donath. "Identity and Deception in the Virtual Community". In: *Communities in Cyberspace*. Ed. by Marc A. Smith and Peter Kollock. London: Routledge, 1999, pp. 29–59.

[76] Erik Moeller. *Response to Larry Sanger's departure from Wikipedia*. Mar. 2002.

[77] AnonymousDonor and Stephen Gilbert. *Larry Sanger (Meatball Wiki discussion of Sanger's departure)*. Mar. 2002.

[78] Wikipedia contributors. *Citizendium*. 2026.

[79] The Wikipedia Signpost. *In the news: Calacanis reignites Sanger founder debate*. Oct. 2009.

[80] Wikipedia contributors. *Transcript for Jimmy Wales: Wikipedia | Lex Fridman Podcast #385 - Lex Fridman*. 2026.

[81] Storyboard18. *Wikipedia's Jimmy Wales Ends Interview in Under a Minute After Dispute over 'Founder' Label*. Network18 Group. Nov. 2025.

[82] Wikipedia contributors. *Wikipedia:Verifiability*. 2026.

[83] Wikipedia contributors. *Larry Sanger*. 2026.

[84] Larry Sanger. *Nine Theses on Wikipedia*. Sept. 2025.

[85] Elon Musk. *"Some good suggestions from the co-founder of Wikipedia"*. Sept. 2025.

[86] Will Oremus. *He co-founded Wikipedia. Now he's inspiring Elon Musk to build a rival*. Oct. 2025.

[87] Larry Sanger. *Grokipedia: a first look*. Oct. 2025.

[88] Emily Dreibelbis Forlini. *Wikipedia's Most Outspoken Critic Says Grokipedia Is Full of AI 'Bullshittery'*. 2025.

[89] Jimmy Wales. *Comment on the proposed Sanger ban, User talk:Jimbo Wales*. 2026.

[90] Jimmy Wales. *[WikiEN-l] Original research*. Dec. 2004.

[91] Wikipedia contributors. *The Evolution of Wikipedia's Norm Network*. 2026.

[92] Jimmy Wales. *Re: [Wikipedia-l] Consensus*. Feb. 2002.

[93] Wikipedia contributors. *Wikipedia:Size of Wikipedia*. 2026.

[94] Wikipedia contributors. *Wikipedia talk:Neutral point of view/Archive 1*. 2026.

[95] Jimmy Wales. *Re: [WikiEN-l] Articles about ourselves*. Nov. 2003.

[96] Wikipedia contributors. *Wikipedia:Neutral point of view*. 2026.

[97] Wikipedia contributors. *Wikipedia:No original research*. 2026.

[98] Heather Ford. *Writing the Revolution: Wikipedia and the Survival of Facts in the Digital Age*. Cambridge, MA: The MIT Press, 2022.

[99] Wikipedia contributors. *Arbitration Committee (Wikipedia)*. 2026.

[100] Wikimedia Foundation. *Memory:Timeline/2003*. 2026.

[101] Jim Giles. "Internet encyclopaedias go head to head". In: *Nature* 438.7070 (Dec. 2005), pp. 900–901.

[102] Encyclopaedia Britannica. *Fatally Flawed: Refuting the Recent Study on Encyclopedic Accuracy by the Journal Nature*. Mar. 2006.

[103] Wikipedia contributors. *Wikipedia:Five pillars*. 2026.

[104] Free Software Foundation. *What is Free Software? — The Free Software Definition*. 2024.

[105] Free Software Foundation. *GNU General Public License, Version 1*. Feb. 1989.

[106] Free Software Foundation. *GNU General Public License, Version 2*. June 1991.

[107] Richard M. Stallman. *Why Free Software Needs Free Documentation*. Free Software Foundation. 2002.

[108] Free Software Foundation. *GNU Free Documentation License, Version 1.1*. Mar. 2000.

[109] Nicholas Carlini et al. "Quantifying Memorization Across Neural Language Models". In: *International Conference on Learning Representations (ICLR)*. 2023.

[110] Mend (formerly WhiteSource). *Open Source Licenses in 2022: Trends and Predictions*. 2022.

[111] Free Software Foundation. *Why the Affero GPL*. 2024.

[112] Wikivoyage contributors. *Why Wikivoyage isn't GFDL*. 2026.

[113] Creative Commons. *History — Creative Commons*. 2024.

[114] GNU Project. *The Free Encyclopedia Project*. 2001.

[115] Free Software Foundation. *GNU Free Documentation License, Version 1.2*. Nov. 2002.

[116] Jonathan Corbet. *GFDL 1.3: Wikipedia's exit permit*. LWN.net. 2008.

[117] Wikipedia contributors. *Creative Commons license*. 2026.

[118] Wikimedia Foundation. *Licensing update/Questions and Answers*. 2009.

[119] Anonymous Slashdot commenter. *Comment on "Wikipedia Moving From GFDL To Creative Commons License"*. May 21, 2009.

[120] *The Architecture of Open Source Applications (Volume 2)MediaWiki*. 2026.

[121] David Gerard. *The FSF is still pushing the GFDL. Ignore them*. May 2011.

[122] Wikipedia contributors. *Wikitravel*. 2026.

[123] Debian Project. *General Resolution: Why the GNU Free Documentation License is not suitable for Debian main*. Mar. 2006.

[124] Wikipedia contributors. *GNU Free Documentation License*. 2026.

[125] Creative Commons. *Wikipedia + CC BY-SA = Free Culture Win!* June 2009.

[126] *An open response to Chris Frey regarding GFDL 1.3 — Free Software Foundation — Working together for free software*. 2026.

[127] Slashdot. *Wikipedia Moving From GFDL To Creative Commons License*. May 2009.

[128] Wikimedia Foundation Board of Trustees. *Resolution: License update*. Dec. 2007.

[129] Wikimedia Foundation Licensing Update Committee. *Licensing update/Result*. Meta-Wiki. May 2009.

[130] Wikipedia contributors. *English Wikipedia*. 2026.

[131] Free Software Foundation. *GNU Free Documentation License, Version 1.3*. Nov. 2008.

[132] Wikimedia Foundation. *Press releases/Dual license vote May 2009*. May 2009.

[133] Erik Moeller. *Licensing resolution*. 2009.

[134] Wikimedia Foundation Board of Trustees. *Resolution: Licensing update approval*. May 2009.

[135] John Seigenthaler. "A false Wikipedia 'biography'". In: *USA Today* (Nov. 2005).

[136] Wikimedia Foundation. *GFDL and CC-BY-SA enforcement - Meta-Wiki*. 2026.

[137] Nature. *Britannica attacks*. Mar. 2006.

[138] Wikimedia Foundation. *Seven years after Nature, pilot study compares*

Wikipedia favorably to other encyclopedias in three languages. Aug. 2012.

[139] The Wikipedia Signpost. *Page creation by anonymous users disabled*. Dec. 2005.

[140] Wikipedia contributors. *Daniel Brandt on the Wikipedia Issue*. 2026.

[141] Wikimedia Foundation. *Terms of Use*. 2024.

[142] Wikimedia, LLC. *Wikimedia API License Agreement*. 2026.

[143] Wikimedia Foundation. *Wikimedia Enterprise: Frequently Asked Questions*. 2021.

[144] Harvard Journal of Law & Technology. *Jacobsen v. Katzer: Federal Circuit Affirms Economic Interest of Open Source Copyright Holder*. 2008.

[145] *License Enforcement - Creative Commons*. 2026.

[146] Software Freedom Conservancy. *Copyleft Compliance Projects*. 2024.

[147] Creative Commons. *Creative Commons Licenses Upheld in Dutch Court*. Mar. 2006.

[148] TidBITS. *Creative Commons License Upheld in Dutch Court*. Mar. 2006.

[149] U.S. District Court for the District of Columbia. *Drauglis v. Kappa Map Group, LLC (No. 1:14-cv-01043; 128 F. Supp. 3d 46)*. 2015.

[150] U.S. District Court for the Eastern District of Virginia. *Philpot v. Media Research Center Inc. (No. 1:17-cv-822)*. Jan. 2018.

[151] *Statement of Enforcement Principles - Creative Commons*. 2026.

[152] Cory Doctorow. *A Bug in Early Creative Commons Licenses Has Enabled a New Breed of Superpredator*. 2021.

[153] Rupert Goodwins. *Automated image recognition: How using 'free' photos on the internet can lead to lawsuits and fines*. 2020.

[154] ABA Journal. *Lawyer deemed 'copyright troll' must file sanctions opinion in all his cases, 2nd Circuit agrees*. 2021.

[155] United States Congress. *Copyright Act of 1976, 17 U.S.C. § 101 (definitions of "copies" and "fixed")*. 1976.

[156] United States Congress. *Copyright Act of 1976, 17 U.S.C. § 106 (exclusive rights in copyrighted works)*. 1976.

[157] U.S. District Court for the Northern District of California. *Bartz v. Anthropic PBC, Order on Fair Use (No. 3:24-cv-05417)*. June 2025.

[158] Skadden, Arps, Slate, Meagher & Flom LLP. *Court Grants Motion To Dismiss in Kadrey AI Training Data Case*. Nov. 2023.

[159] Andres Guadamuz. *Creative Commons and AI training*. 2023.

[160] U.S. Copyright Office. *Copyright and Artificial Intelligence, Part 3: Generative AI Training (Pre-Publication Version)*. May 2025.

[161] Brian Lamb. *Q&A: Jimmy Wales, Wikipedia Founder*. Sept. 2005.

[162] Wikimedia Foundation. *History of the Wikimedia Foundation Board of Trustees*. 2026.

[163] Wikimedia Foundation. *Wikimedia Foundation elections/Board elections/2004 - Meta-Wiki*. 2026.

[164] Wikipedia contributors. *Digital 2025: exploring trends in Wikipedia traffic — DataReportal – Global Digital Insights*. 2026.

[165] Wikimedia Foundation. *Founder's seat - Meta-Wiki*. 2026.

[166] Wikipedia contributors. *Wikipedia:Wikipedia Signpost/2022-10-31/News and notes*. 2026.

[167] Wikipedia contributors. *Wikipedia Endowment: Is Wikipedia Going Broke?* 2026.

[168] Wikipedia contributors. *Protests against SOPA and PIPA*. 2026.

[169] Wikimedia Foundation. *Wikipedia:SOPA initiative/Learn more*. 2012.

[170] Electronic Frontier Foundation. *After Historic Protest, Members of Congress Abandon PIPA and SOPA in Droves*. 2012.

[171] ABC News. *Wikipedia Blackout: Websites Wikipedia, Reddit, Others Go Dark Wednesday to Protest SOPA, PIPA*. 2012.

[172] Wikipedia contributors. *Sue Gardner*. 2026.

[173] Wikimedia Foundation. *Fundraising 2010/Report*. 2011.

[174] The Globe and Mail. *Wikipedians do it for love. Really.* 2009.

[175] Know Your Meme. *2010 Wikipedia Fundraising Campaign*. 2010.

[176] Instant Fundas. *Wikipedia's Jimmy Wales Appeal increases donation by 15x*. 2010.

[177] NBC News. *Jimmy Wales' creepy stare rockets $16 million in Wikipedia donations*. 2012.

[178] Wikimedia Foundation. *Fundraising 2011/Report*. 2012.

[179] Sue Gardner. *Wikipedia's community calls for anti-SOPA blackout January 18*. 2012.

[180] Agence France-Presse. *Wikipedia back online after blackout protest*. 2012.

[181] Amit Singhal. *Introducing the Knowledge Graph: things, not strings*. 2012.

[182] Wikipedia contributors. *Relationship between Google and Wikipedia*. 2026.

[183] Wikipedia contributors. *Wikipedia urges AI companies to use its paid API, and stop scraping | TechCrunch*. 2026.

[184] Voicebot.ai. *Voice Assistants Alexa, Bixby, Google Assistant and Siri Rely on Wikipedia and Yelp to Answer Many Common Questions about Brands*. 2019.

[185] Wikimedia Endowment. *2023–2024 Annual Report*. 2024.

[186] Wikipedia Signpost. *News and notes: Media Viewer, Superprotect, and German Wikipedia*. 2014.

[187] Wikimedia Foundation. *Memory:Timeline*. 2026.

[188] Wikipedia Signpost. *Foundation to relocate to San Francisco*. 2007.

[189] Sue Gardner. *Please read: an announcement from Wikimedia Foundation ED Sue Gardner*. 2013.

[190] Wikimedia Foundation. *Wikimedia Foundation's Sue Gardner named to World's Most Powerful Women list by Forbes*. 2012.

[191] English Wikipedia. *Wikimedia Foundation*. 2026.

[192] Wikimedia Foundation. *Research:The Rise and Decline - Meta-Wiki*. 2026.

[193] Wikimedia Foundation. *Press releases/Stanton Foundation Awards Wikimedia $3.6 Million for Technology Improvements*. 2011.

[194] Wikimedia Foundation. *VisualEditor beta rollout*. 2013.

[195] English Wikipedia. *VisualEditor/Default State RFC*. 2013.

[196] Wikipedia Signpost. *News and notes: VisualEditor now opt-in only on English Wikipedia*. 2013.

[197] Meta-Wiki. *Superprotect*. 2015.

[198] Meta-Wiki. *Letter to Wikimedia Foundation: Superprotect and Media Viewer*. 2014.

[199] Meta-Wiki. *Requests for comment/Superprotect rights*. 2014.

[200] Wikipedia Signpost. *News and notes: Superprotect removed*. 2015.

[201] MediaWiki. *WMF product development process/2015-11-05*. 2015.

[202] Wikipedia contributors. *Wikipedia:Wikipedia Signpost/2016-02-17/Special report*. 2026.

[203] Wikimedia Foundation. *Wikimedia Foundation Board noticeboard/James Heilman removal FAQ - Meta-Wiki*. 2026.

[204] Wikimedia Foundation. *Resolution:James Heilman Removal*. 2026.

[205] Wikimedia Foundation. *Account of the events leading to James Heilman's removal*. 2026.

[206] Wikipedia Signpost. *Special report: Knowledge Engine revealed*. 2016.

[207] English Wikipedia. *Community response to the Wikimedia Foundation's ban of Fram*. 2019.

[208] Avi Asher-Schapiro. *Wikipedia Is in a Constitutional Crisis Over the Banning of an Editor Named Fram*. 2019.

[209] Wikipedia Signpost. *Special report: Fewer than 500 active admins*. 2019.

[210] Wikipedia Signpost. *Special report: ArbCom unbans Fram*. 2019.

[211] Wikimedia Foundation. *The Wikimedia Endowment reaches $100 million milestone and welcomes three new members to its Board: more on what these developments*

mean for the projects and movement – Diff. 2026.

[212] Guy Macon. *Wikipedia has Cancer*. 2017.

[213] Wikimedia Foundation. *Return of Organization Exempt From Income Tax (Form 990), Fiscal Year 2023–2024*. 2025.

[214] Goran Penić. *Jovanović pozvao učenike i studente na bojkot hr.wiki 'Sadržaj je falsificiran'*. Jutarnji list. Sept. 13, 2013.

[215] Tihana Tomičić. *Jovanović: Djeco, ne baratajte hrvatskom Wikipedijom jer su sadržaji falsificirani*. Novi list. Sept. 13, 2013.

[216] Croatian Wikipedia contributors. *Wikipedija na hrvatskome jeziku*.

[217] Wikimedia Foundation Trust and Safety Disinformation Team. *Croatian WP Disinformation Assessment - Final Report (EN)*. June 14, 2021.

[218] Lovro Krnić. *Endehapedia*. Portal Novosti. May 28, 2018.

[219] MJL and Meta-Wiki contributors. *Requests for comment/Global ban for Kubura*. Nov. 28, 2020.

[220] Wikimedia Foundation. *Croatian Wikipedia Disinformation Assessment-2021*. June 14, 2021.

[221] Sven Milekic. *How Croatian Wikipedia Made a Concentration Camp Disappear*. Balkan Insight. Mar. 26, 2018.

[222] Out Magazine. *Croatian Official Slams Anti-Gay Wikipedia Edits*. Sept. 17, 2013.

[223] Lasta and Meta-Wiki stewards. *Steward requests/Checkuser/2020-11*. Nov. 2020.

[224] Goran Penić. *'NDH nije bila totalitarna, a žrtve u Jasenovcu pobili su partizani': Desničari preuzeli uređivanje hrvatske Wikipedije*. Jutarnji list. Sept. 10, 2013.

[225] Meta-Wiki. *Requests for comment/2013 issues on Croatian Wikipedia*. Sept. 13, 2013.

[226] Wikipedia Signpost. *The curious case of the Croatian Wikipedia*. Aug. 30, 2019.

[227] Wikipedia Signpost. *Disinformation report*. June 27, 2021.

[228] Zarine Kharazian, Kate Starbird, and Benjamin Mako Hill. "Governance Capture in a Self-Governing Community: A Qualitative Comparison of the Serbo-Croatian Wikipedias". In: *Proceedings of the ACM on Human-Computer Interaction* 8.CSCW1 (2024).

[229] Simon Owens. *The battle to destroy Wikipedia's biggest sockpuppet army*. Oct. 2013.

[230] Sue Gardner. *Wikimedia Foundation Executive Director Sue Gardner's response to paid advocacy editing and sockpuppetry*. Oct. 2013.

[231] Charles Arthur. *Wikipedia sends cease-and-desist letter to PR firm offering paid edits to site*. Nov. 2013.

[232] Stephen LaPorte, Luis Villa, and Geoff Brigham. *Making a change to our Terms of Use: Requirements for disclosure*. June 2014.

[233] Wikimedia Foundation. *Hundreds of "black hat" English Wikipedia accounts blocked following investigation*. Aug. 2015.

[234] *Massive paid editing network unearthed on the English Wikipedia: Nearly 400 accounts blocked in largest paid-editing bust ever*. Sept. 2015.

[235] Jamie Merrill. *Wikipedia 'rogue editors' have targeted hundreds of people in a blackmail scam*. Sept. 2015.

[236] Wikimedia Foundation. *Wikimedia Foundation reports/Financial/Audits/2019-2020 – frequently asked questions*. Oct. 2020.

[237] Meta-Wiki. *Knowledge Equity Fund/Frequently asked questions*.

[238] Wikimedia Foundation. *Wikimedia Foundation Announces First Grant Recipients of New $4.5 Million Equity Fund to Close Knowledge Gaps and Promote Racial Equity*. Sept. 8, 2021.

[239] Yair Rand. *WMF transfers $8.7 million to "Wikimedia Knowledge Equity Fund"*. Dec. 13, 2020.

[240] Wikipedia Signpost. *News and notes: Knowledge Equity Fund controversy*. Oct. 31, 2022.

[241] Wikimedia Foundation. *What We've Learned From the Equity Fund's First Round*. Apr. 12, 2023.

[242] Wikipedia Signpost. *News and notes: Village Pump RfC on Knowledge Equity Fund mission creep (35:23)*. Oct. 23, 2023.

[243] Wikipedia Signpost. *News and notes: Knowledge Equity Fund final round and closure*. Oct. 20, 2025.

[244] Wikidata. *Wikimedia, LLC (Q106204654)*. 2024.

[245] Noam Cohen. "Wikipedia Is Finally Asking Big Tech to Pay Up". In: *Wired* (Mar. 2021).

[246] Wikimedia Foundation. *Fundraising - Meta-Wiki*. 2026.

[247] Wikipedia contributors. *Are corporations that use Wikipedia giving back? | TechCrunch*. 2026.

[248] Wikipedia contributors. *Amazon Alexa*. 2026.

[249] Wikipedia contributors. *New User Trends on Wikipedia – Diff*. 2026.

[250] *Google AI Overviews drive 61% drop in organic CTR, 68% in paid*. 2026.

[251] Amit Kapoor. *Wikipedia seeks global operator partners to enable free access*. Oct. 2011.

[252] Amit Kapoor. *Marketing Free Knowledge on Mobile in Africa*. June 2012.

[253] Wikimedia Foundation. *Building for the future of Wikimedia with a new approach to partnerships*. Feb. 2018.

[254] Electronic Frontier Foundation. *Net Neutrality and the Global Digital Divide*. July 2014.

[255] Access Now. *Wikipedia Zero and net neutrality: Wikimedia turns its back on the open internet*. Aug. 2014.

[256] Brian Fung. *Wikipedia's 'complicated' relationship with net neutrality*. Nov. 2014.

[257] Telecom Regulatory Authority of India. *Prohibition of Discriminatory Tariffs for Data Services Regulations, 2016*. Feb. 2016.

[258] Aakriti Srivastava and Nikhil Pahwa. *Wikipedia Zero to be discontinued; Facebook's FreeBasics should follow suit*. Feb. 2018.

[259] Intersection Group. *Lane Becker*. 2025.

[260] Wikimedia Foundation. *Wikimedia Foundation launches Wikimedia Enterprise, the new, opt-in product for companies and organizations to easily reuse content from Wikipedia and Wikimedia projects*. Oct. 2021.

[261] Wikimedia Enterprise. *Structured Contents*. 2024.

[262] Wikimedia Foundation. *Wikimedia Enterprise announces Google and Internet Archive as its first customers*. June 2022.

[263] Wikimedia Foundation. *How crawlers impact the operations of the Wikimedia projects – Diff*. 2026.

[264] Bloomberg. *Wikipedia Co-Founder: AI Bots Costing Us a Lot of Money*. Dec. 2025.

[265] Associated Press. *Wikipedia inks AI deals with Microsoft, Meta and Perplexity as it marks 25th birthday*. Jan. 2026.

[266] James Boyle. "The Second Enclosure Movement and the Construction of the Public Domain". In: *Law and Contemporary Problems* 66.1 (2003), pp. 33–74.

[267] OpenAI. *GPT-4 Technical Report*. 2023.

[268] Euronews. *OpenAI says it's 'impossible' to train AI without copyrighted materials*. Jan. 2024.

[269] RankStudio. *What Is Common Crawl? A History of the Open Web Dataset*. 2024.

[270] Stefan Baack. *Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI*. Feb. 2024.

[271] Ahmed Ahmed et al. *Extracting Books from Production Language Models*. Jan. 2026.

[272] Athena Chapekis, Colleen McClain, and Michael Lipka. *Google users are less likely to click on links when an AI summary appears in the results*. July 2025.

[273] U.S. Court of Appeals, Third Circuit. *Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240 (3d Cir.)* 1983.

[274] U.S. Court of Appeals, Ninth Circuit. *MAI Systems Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir.)* 1993.

[275] John X. Morris et al. *How Much Do Language Models Memorize?* 2025.

[276] Stella Biderman et al. *Emergent and Predictable Memorization in Large Language Models*. 2023.

[277] Vitaly Feldman. *Does Learning Require Memorization? A Short Tale about a Long Tail*. 2020.

[278] OpenAI. *Accelerating the Next Phase of AI*. Mar. 2026.

[279] CNBC. *OpenAI Closes Funding Round at an $852 Billion Valuation*. Mar. 2026.

[280] Anthropic. *Anthropic Raises $65B in Series H Funding*. May 2026.

[281] TechCrunch. *Anthropic Raises $65 Billion, Nears $1T Valuation Ahead of IPO*. May 2026.

[282] OpenAI. *Introducing deep research*. Feb. 2025.

[283] Wikimedia Research Newsletter. *Recent research: Will ChatGPT diminish Wikipedia? (March 2025)*. Mar. 2025.

[284] Noam Brown. *Twitter thread on Deep Research and Wikipedia*. Feb. 2025.

[285] Christian Wagner and Ling Jiang. "Death by AI: Will large language models diminish Wikipedia?" In: *Journal of the Association for Information Science and Technology* 76.5 (2025), pp. 743–751.

[286] Reuters. *OpenAI CFO says annualized revenue crosses $20 billion in 2025*. 2025.

[287] OpenAI. *Introducing ChatGPT*. Nov. 2022.

[288] Wikipedia contributors. *Can humans and bots share the Internet? Wikipedia thinks so. | IBM*. 2026.

[289] 2026.

[290] Fairly Trained. *Announcing Our Newly Certified Models and New Supporters*. Mar. 2024.

[291] Fairly Trained. *Licensed Model (L) Certification*. 2024.

[292] U.S. District Court, District of Delaware. *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc*. 2025.

[293] Mia Sato. "How Jeffrey Epstein used SEO to bury news about his crimes". In: *The Verge* (Nov. 2025).

[294] Wikipedia contributors. *Wikipedia:Wikipedia Signpost/2020-03-29/In focus: Jeffrey Epstein has asked me...* 2020.

[295] *Al seckel: Results — Epstein Emails*. 2025.

[296] Wikipedia contributors. *Al Seckel*. 2026.

[297] *Al seckel: Re: Results — Epstein Emails*. 2025.

[298] Dflovett. *Did Jeffrey Epstein edit his own Wikipedia article?* 2025.

[299] George J. Stigler. "The Theory of Economic Regulation". In: *The Bell Journal of Economics and*

Management Science 2.1 (1971), pp. 3–21.

[300] Wikipedia. *Katherine Maher*. 2026.

[301] Wikipedia contributors. *Jimmy Wales*. 2026.

[302] Wikipedia. *Maryana Iskander*. 2026.

[303] Wikipedia contributors. *I Founded Wikipedia. Here's How to Fix It.* 2026.

[304] NBC News. *Elon Musk launches Grokipedia as an alternative to 'woke' Wikipedia*. 2025.

[305] Wikipedia contributors. *Wikipedia:Reliable sources/Perennial sources*. 2026.

[306] NPR. *Grokipedia and Wikipedia: Here's what they say about each other*. 2025.

[307] PolitiFact. *Musk's AI-powered Grokipedia: A Wikipedia spin-off with less care to sourcing, accuracy*. 2025.

[308] Fox Business. *Elon Musk launches Grokipedia, AI rival to Wikipedia with 885K articles*. 2025.

[309] Siladitya Ray. *Musk Launches 'Grokipedia'—An AI-Generated Wikipedia Challenger*. 2025.

[310] The Verge. *Elon Musk's Grokipedia launches with AI-cloned pages from Wikipedia*. 2025.

[311] Poynter. *Grokipedia touts AI-powered depth. The reality: heavily borrowed Wikipedia entries*. 2025.

[312] Tekendra Parmar. *Elon Musk's Anti-Woke Wikipedia Is Calling Hitler "The Führer"*. 2025.

[313] OpenAI. *ChatGPT Pricing*. 2026.

[314] CNBC. *Elon Musk's Grokipedia is live. Wikipedia's founder isn't impressed*. 2025.

[315] Wikipedia contributors. *Wikipedia announces AI deals with Amazon, Meta, Perplexity and more*. 2026.

[316] Wikipedia contributors. *Why Wikipedia Must Jettison Its Anti-Elitism – LarrySanger.org*. 2026.

[317] Straight Arrow News. *Wikipedia co-founder reacts to Musk's Grokipedia announcement*. 2025.

[318] Larry Sanger. *My initial impression … is that Grokipedia is very OK*. 2025.

[319] DemandSage. *Grokipedia Statistics (2026): Active Users & Traffic*. 2026.

[320] Engadget. *Report reveals that OpenAI's GPT-5.2 model cites Grokipedia*. 2026.

[321] The Register. *Grokipedia is such a mess even Grok thinks its untrustworthy*. 2025.

[322] The Guardian. *In Grok we don't trust: academics assess Elon Musk's AI-powered encyclopedia*. 2025.

[323] Wikipedia contributors. *Irving v Penguin Books Ltd*. 2026.

[324] Richard J. Evans. *Lying About Hitler: History, Holocaust, and the David Irving Trial*. Basic Books, 2001.

[325] OpenAI. *Introducing GPT-5.2*. 2025.

[326] Ilia Shumailov et al. "AI models collapse when trained on recursively generated data". In: *Nature* 631 (2024), pp. 755–759.

[327] Silvia Williams and Rice University. *Breaking MAD: Generative AI could break the internet*. 2024.

[328] Matthias Gerstgrasser, Rylan Schaeffer, et al. *Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data*. 2024.

[329] Xibeijia Guan. *What percentage of new content is AI-generated? (Study of 900k pages)*. 2025.

[330] Originality.ai. *Amount of AI Content in Google Search Results — Ongoing Study*. 2025.

[331] NewsGuard. *Tracking AI-enabled Misinformation: AI Tracking Center*. 2025.

[332] Europol Innovation Lab. *Facing reality? Law enforcement and the challenge of deepfakes*. 2022.

[333] OpenAI. *OpenAI o3 and o4-mini System Card*. 2025.

[334] Wikipedia contributors. *Citogenesis: the serious circular reporting problem Wikipedians are fighting*. 2026.

[335] Shane Greenstein and Feng Zhu. "Is Wikipedia Biased?" In: *American Economic Review* 102.3 (2012), pp. 343–348.

[336] Anti-Defamation League. *New ADL Report Finds Evidence of Biased, Coordinated Campaign on Wikipedia Related to the Israeli–Palestinian Conflict*. Mar. 2025.

[337] Jan Grabowski and Shira Klein. "Wikipedia's Intentional Distortion of the History of the Holocaust". In: *The Journal of Holocaust Research* 37.2 (2023), pp. 133–190.

[338] Wikimedia Foundation. *Croatian Wikipedia Disinformation Assessment*. 2021.

[339] American Sunlight Project. *New report: Russian propaganda may be flooding AI models*. 2025.

[340] EU DisinfoLab. *LLM grooming: a new strategy to weaponise AI for FIMI purposes*. 2025.

[341] Wikimedia Foundation. *Meet the Wikimedians of the Year 2025*. 2025.

[342] Wikipedia contributors. *The Editors Protecting Wikipedia from AI Hoaxes*. 2026.

[343] Wikipedia contributors. *Wikipedia:WikiProject AI Cleanup*. 2026.

[344] Wikipedia contributors. *Wikipedia:Signs of AI writing*. 2026.

[345] Wikipedia contributors. *The Rise of AI-Generated Content in Wikipedia*. 2026.

[346] Wikipedia contributors. *Wikipedia:Wikipedia Signpost/About*. 2026.

[347] Matthew A. Vetter. *Wikipedia at 25: Can Its Original Ideals Survive in the Age of AI?* 2026.

[348] Christophe Henner. *Wikipedia at 25: A Wake-Up Call*. 2026.

[349] Wikipedia contributors. *Wikipedia says it will use AI, but not to replace human volunteers | TechCrunch*. 2026.

[350] Wikipedia contributors. *Wikipedia:Articles for creation*. 2026.

[351] Wikipedia contributors. *Wikipedia:Snowball clause*. 2026.

[352] The Signpost. *Wikipedia:Wikipedia Signpost/Single/2025-10-20*. 2025.

[353] Wikipedia contributors. *Wikipedia:What Wikipedia is not (Wikipedia is not a battleground)*. 2026.

[354] Wikipedia contributors. *Wikipedia:Canvassing*. 2026.

[355] Meta-Wiki. *Movement Charter/Ratification/Voting/Results*. 2024.

[356] Wikimedia Foundation Board of Trustees. *Resolution: Draft Movement Charter and next steps*. July 8, 2024.

[357] Charter Electoral Commission. *Movement Charter/Drafting Committee/Announcement — Results of the ratification vote*. 2024.

[358] The Wikipedia Signpost. *News and notes: Movement Charter not ratified by the Wikimedia Foundation*. July 22, 2024.

[359] Nataliia Tymkiv. *Wikimedia Foundation Board resolution and vote on the proposed Movement Charter*. July 11, 2024.

[360] Wikimedia Foundation Board of Trustees. *Wikimedia Foundation Board noticeboard/October 2025 update*. Oct. 3, 2025.

[361] Hannah Clover. *2025 WMF Board reform petition*. Oct. 9, 2025.

[362] Arabic Wikimedia community. *Objections to the 2025 WMF Board election removals/Arab Community*. Oct. 2025.

[363] Lane Rasberry. *Lane not on ballot for WMF board of trustees (Wikipedia talk:Wikipedia Signpost)*. Oct. 3, 2025.

[364] The Wikipedia Signpost. *Special report: the 2025 WMF Board of Trustees election controversy*. Oct. 20, 2025.

[365] Wikipedia contributors. *Wikipedia:About*. 2026.

[366] Wikipedia contributors. *Wikipedia:Size of Wikipedia*. 2026.

[367] Christine Hall. *Seven Years After, Stallman Is Still Stallman*. 2026.

[368] Slashdot. *Richard Stallman Critiques AI, Connected Cars, Smartphones, and DRM*. 2026.

[369] *FSF is working on freedom in machine learning applications — Free Software Foundation — Working together for free software*. 2026.

[370] FOSDEM. *FSF's criteria for free machine learning applications*. 2025.

[371] Sara Fischer. *OpenAI says ChatGPT usage has doubled in the last year*. 2024.

[372] Creative Commons. *Understanding CC Licenses and AI Training: A Legal Primer*. May 2025.

[373] Patronus AI. *Introducing CopyrightCatcher, the first Copyright Detection API for LLMs*. Mar. 2024.

[374] Cornell Chronicle. *ChatGPT 'memorizes' and spits out entire poems*. Jan. 2024.

[375] U.S. Supreme Court. *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*. 2023.

[376] U.S. Supreme Court. *Feist Publications, Inc. v. Rural Telephone Service Co.* 1991.

[377] U.S. District Court, Northern District of California. *Kadrey v. Meta Platforms, Inc., Order on Cross-Motions for Summary Judgment*. June 2025.

[378] Search Engine Roundtable. *Similarweb Says No Clicks From Google Grew From 56% to 69% Since AI Overviews*. 2025.

[379] Martin Senftleben. *The Bartz v. Anthropic Settlement: Understanding America's Largest Copyright Settlement*. 2025.

[380] Authors Guild. *What Authors Need to Know About the $1.5 Billion Anthropic Settlement*. 2026.

[381] The New York Times Company. *The New York Times Company v. Microsoft Corporation and OpenAI*. Dec. 2023.

[382] NPR. *Judge allows 'New York Times' copyright case against OpenAI to go forward*. Mar. 2025.

[383] AI Lawsuit Tracker. *New York Times v. OpenAI & Microsoft*. 2026.

[384] National Law Review. *OpenAI Loses Privacy Gambit: 20 Million ChatGPT Logs Likely Headed to Copyright Plaintiffs*. 2026.

[385] Campaign US. *1% click-through rate: How Google AI Overviews is killing publishers*. 2025.

[386] Wikimedia Foundation. *Policy: Wikimedia Foundation Commercial Sales and Contracts Policy*. May 2025.

[387] Futurism. *After Being Pillaged By AI Companies, Wikipedia Signs Deal to Get Paid By Them*. Jan. 2026.

[388] Michelle Paulson and Geoff Brigham. *Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance*. Wikimedia Foundation. Mar. 10, 2015.

[389] American Civil Liberties Union. *Wikimedia v. NSA — Challenge to Upstream Surveillance*. American Civil Liberties Union.

[390] Jimmy Wales and Lila Tretikov. *Stop Spying on Wikipedia Users*. Mar. 10, 2015.

[391] The Signpost. *In focus: Wikimedia sues the NSA*. Wikipedia Signpost. Mar. 11, 2015.

[392] Jim Buatti and Aeryn Palmer. *Victory at the Fourth Circuit: Court*

of Appeals allows Wikimedia Foundation v. NSA to proceed. Wikimedia Foundation. May 23, 2017.

[393] Wikimedia Foundation. *Wikimedia Foundation, ACLU, and Knight Institute Urge U.S. Supreme Court to Hear Challenge to NSA's Mass Surveillance*. Wikimedia Foundation. Aug. 26, 2022.

[394] "U.S. Supreme Court Declines to Hear Wikimedia Foundation's Challenge to NSA Mass Surveillance". In: *Wikimedia Diff* (Feb. 2023).

[395] Louise Stewart. *Wikimedia Says When a Monkey Takes a Selfie, No One Owns It*. Aug. 21, 2014.

[396] Wikimedia Foundation. *Mailbag: What is going on with Wikipedia and the monkey selfie?* Wikimedia Foundation. Dec. 22, 2017.

[397] BBC News. *Photographer 'lost £10,000' in Wikipedia monkey 'selfie' row*. BBC News. Aug. 7, 2014.

[398] U.S. Copyright Office. *Compendium of U.S. Copyright Office Practices, 3d ed., §313.2*. United States Copyright Office. Dec. 22, 2014.

[399] Chat GPT Is Eating the World. *Latest US Map of Copyright Suits v. AI Companies*. June 2026.

[400] Bloomberg Law. *OpenAI Copyright Suit Consolidation Portends Consistency, Risk*. 2025.

[401] Wikimedia Foundation. *Meet the Wikimedians of the Year 2025*. 2025.

[402] Oded Nov. "What Motivates Wikipedians?" In: *Communications of the ACM* 50.11 (2007), pp. 60–64.

[403] Michael Restivo and Arnout van de Rijt. "Experimental Study of Informal Rewards in Peer Production". In: *PLOS ONE* 7.3 (2012), e34358.

[404] Christophe Henner. *Wikipedia at 25: A Wake-Up Call*. 2026.

[405] Wikipedia contributors. *Wikipedia:Essays*. 2026.

[406] *Don't bite the newbies | Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. 2026.

[407] Wikimedia Foundation. *Contributors/Mobile Web Editing Research*. 2025.

[408] Sherwood News. *ChatGPT Is Officially Clocking More Site Visits Than Wikipedia in the US*. 2025.

[409] Barry Schwartz. *Similarweb Says No Clicks From Google Grew From 56% to 69% Since AI Overviews*. 2025.

[410] Ben Brubaker. *At 25, Wikipedia Now Faces Its Most Existential Threat—Generative A.I.* 2026.

[411] Christian Wagner and Ling Jiang. "Death by AI: Will Large Language Models Diminish Wikipedia?" In: *Journal of the Association for Information Science and Technology* 76.5 (2025), pp. 743–751.

[412] Blake Brittain and IPWatchdog. *ROSS Intelligence Appeals Originality, Fair Use Rulings in*

Thomson Reuters AI Legal Tool Lawsuit. Apr. 2025.

[413] Jonas Gamalielsson and Björn Lundell. "Long-Term Sustainability of Open Source Software Communities beyond a Fork: A Case Study of LibreOffice". In: *Open Source Systems: Long-Term Sustainability (OSS 2012)*. Vol. 378. IFIP Advances in Information and Communication Technology. Springer, 2012, pp. 29–47.

[414] U.S. Court of Appeals, Third Circuit. *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc., No. 25-2153 (docket)*. 2026.

[415] The Nobel Foundation. *Elinor Ostrom – Facts*. 2009.

[416] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Political Economy of Institutions and Decisions. Cambridge University Press, 1990.

[417] Mend.io. *Open Source Licenses in 2022: Trends and Predictions*. 2022.

[418] Jono Bacon. *The decline of GPL?* 2017.

[419] Wikimedia Foundation. *Wikimedia Foundation reports/Financial/Audits/2024–2025 — frequently asked questions*. 2025.

Index

Colophon

This book was designed and typeset by the author with LuaLaTeX, at eleven point on a six-by-nine-inch trim. Every edition—the print interior, the wraparound cover, and the ebook—builds from the same source files, kept under version control from first outline to final page.

The text is set in Libertinus Serif, the freely licensed successor to Linux Libertine: the typeface of the Wikipedia wordmark itself. A book about the encyclopedia's commons, set in the letters the encyclopedia wears on its own logo, and licensed to remain free the same way. Code and URLs are set in Latin Modern Mono. The ebook edition is set in EB Garamond and produced by a custom LaTeX-to-EPUB converter written for this project.

The covers were drawn in TikZ. The cover art is the author's derivative of the Wikipedia puzzle globe, and carries its CC BY-SA license forward, as share-alike requires. Citations are managed with biblatex and biber; references to Wikipedia's own pages cite pinned revisions, so the version quoted is the version cited.

Like the commons it describes, this book is licensed CC BY-SA 4.0. Use it freely, but keep it free.

Michigan · 2026 · commonscaptured.com