
The Never-Ending Library

When the FBI seized the online Z-Library in 2022, it looked like the final chapter for free, illegal book-sharing. Instead, its legacy has become central to the AI revolution

BY FELIX BAZALGETTE
ILLUSTRATIONS BY MUHAMMAD FATCHUROFI



The idea of a total library, one containing all the books in existence, had a peculiar hold on the mind of the Argentine writer Jorge Luis Borges. In 1939, he wrote an essay on the subject, tracing its history from the ancient philosopher Democritus to the 19th-century physicist Gustav Theodor Fechner, and then a short story imagining this library's architecture and the psychological effect it would have on humans, were it ever finished. An initial reaction of "unbounded joy", he thought, would curdle into a "disproportionate depression".

Borges, himself a librarian by profession and always drawn to open-ended lists, mused on the works that would lie in this total library, including: "the songs the sirens sang", "my dreams and half-dreams at dawn on August 14, 1934", "the secret and true name of Rome", "the complete catalog of the Library" and "the proof of the inaccuracy of that catalog".

Were Borges writing today, he might include another text - a sealed complaint filed by the US Justice Department in a federal court in New York on October 21 2022. It was because of this document that 33-year-old Anton Napolsky and 27-year-old Valeriia Ermakova, two librarians travelling to Borges' native country, unknowingly tripped an Interpol Red Notice when they arrived at an airport near Córdoba, in central Argentina, on October 22. Looking forward to two weeks' holiday, and unaware of the bureaucratic storm gathering above them, they rented a car and drove south towards the picturesque lakeside town of El Calafate. Argentine authorities tracked them via their phones.

The complaint, now public, accused the two Russians of running the largest illegal library in the world, a website with tens of millions of digital books, and millions of regular users, called Z-Library. The site provided an easily searchable database, enabling anyone to freely download pirated books, and also solicited users to upload their own files to its ever-expanding collection.

When the pair reached the southern city of Rio Gallegos on November 3, officers from the Argentine Special Operations Unit arrested them. At the same time, all around the world, anyone logging on to Z-Library encountered a stark graphic stamped with the seal of the US Justice Department: *This domain has been seized by the FBI*.

"It's never nice to see a library burn down," wrote Balázs Bodó, a professor of information law at the University of Amsterdam, "and this is what burning down a library looks like in the digital space."

To their supporters and their users, organisations such as Z-Library are known not as book piracy websites but as "shadow libraries", and Bodó has written extensively on their history. According to him, this is not a story of crime and punishment - Russians illegally profiting from stolen goods - but rather a confrontation between two fundamental ideas of how the internet should work, perhaps even how the world should work.

Since Napolsky and Ermakova were arrested, shadow libraries have moved from the margins - quietly used by readers and researchers - to the curious centre of the AI boom. Recent court cases have shown that some of the wealthiest tech companies in the world have relied on millions



Are they right to say that western-enforced copyright has become unmoored from its original purposes, no longer benefiting authors?

of books from shadow libraries for the training of large language models (LLMs). In their ravenous quest for the written word, these companies have devoted vast resources towards achieving what seemed, to Borges, only a fable: a library of everything ever written.

Last year, I met with Bodó, a youthful, slightly twitchy 51-year-old. At his suggestion, we had lunch in the café of Amsterdam's oldest zoo. Bodó wore a pair of tiny round metal-frame glasses and a Hawaiian shirt. He told me about his interest in the underground transmission of knowledge, sparked by his time travelling in Bosnia in the weeks after the war ended in 1995. Back then he noticed that pirate media - bootleg VHS and cassette tapes - were the first to re-enter the country, well before any official distribution companies would tackle such a disrupted and risky market. "In the midst

of the chaos," he told me, describing joyous queues at pirate market stalls, "it was the pirates who were helping Bosnians back into contact with the world, with the wider culture."

Bodó grew up in Hungary, which was, for the first 15 years of his life, a communist country. Almost all of the major book shadow libraries in operation today are founded and run by people who grew up in Russia or post-Soviet satellite states. This, according to Bodó and to the pirates themselves, is not a coincidence.

"The Soviet tradition of special respect for libraries, books and education in general largely shaped our project's ideology," one longtime administrator for Z-Library told me recently. Bodó quotes the Russian sociologist Valeria Stelmakh, who has argued that in the USSR "reading almost transplanted religion as a sacred activity" and that "literature

became the unique source of moral truth for the population”. Bodó points to the Soviet-era *samizdat* traditions of private copying, hoarding and sharing of texts in an environment of official censorship, alongside “the economic ruin of the intelligentsia in the post-communist period”.

Pirates argue that the copyright regime, guarded internationally by the US Justice Department and publishers’ associations in the US and the UK, is a recent, western and capitalist form of censorship, imposed upon the rest of the world. “A tool for profit extraction,” said the Z-Library administrator, “against the public interest, and usually to the benefit of publishers not authors”. They argued that Z-Library’s very existence “demonstrates systemic problems with information access in modern society”.

Intrigued by these grand claims, and taken aback by the vehemence with which the FBI would hunt down a book-sharing website, I set out to understand the world of shadow libraries, talking to publishers’ associations, academics like Bodó, writers, researchers and the pirates themselves.

Where do these organisations come from and how do they work? Are they communist folk heroes or dissembling crypto-capitalist grifters? Are they right to say that western-enforced copyright has become unmoored from its original purposes, no longer benefiting authors? And, with Z-Library back online, where are those internationally wanted librarians today?

DURING THE SECOND WORLD WAR, WHEN Borges first published his writings on *la biblioteca total*, the notion of a total library entered the realm of possibility thanks to advances in information technology. In 1945, in The Atlantic, a government scientist called Vannevar Bush proposed a kind of proto-internet, using networked forms of microphotography that were able to shrink down physical bookshelves. “A library of a million volumes could be compressed into one end of a desk,” he wrote. He speculated that the total sum of human written production – “a billion books, the whole affair” – would soon easily be “lugged off in a moving van”.

The actual practice of digital librarianship is almost as old as the internet itself – Project Gutenberg, an attempt to make digital versions of public-domain texts available to all, was first launched on the Arpanet, a US government-backed precursor to the internet, in 1971. Its first text, typed out manually by a student at the University of Illinois, was the Declaration of Independence, followed by the Constitution and then the King James Bible, which was finally finished in 1989.

The proliferation of scanning technology, leaps in digital storage efficiency and the rise of the World Wide Web in the 1990s led to an explosion in book digitisation in the early 2000s. Now books did not have to be laboriously typed out, and could be stored in online databases accessible from anywhere. Some projects, such as Google’s Project Ocean – 25 million books scanned and digitised between 2002 and 2015 – were run by corporations in partnership with universities. These were subject to early lawsuits from publishers’ and authors’ associations, which limited how they could be accessed and used by the public.

At the same time, a number of smaller, grassroots digitisation projects were emerging, operating outside of the legal regimes that Google was

ultimately subject to. Though they were likened to file-sharing pirate websites such as Napster and The Pirate Bay, which primarily dealt in music or movie files, many of the text-sharing projects consciously eschewed the “pirate” label. They preferred the organising principles, history and brand of the “library”. These projects sought to amass as many digital books as possible – through crowdsourced scanning or the amalgamation of existing databases – catalogue and index them, and make them freely accessible.

One such project, Library Genesis or LibGen, was started by Russian scientists in 2008 to enable the free circulation of scientific papers. Another, Library.nu – founded in the mid-noughties by students operating out of Galway, according to The Sunday Times – became so large that it attracted the attention of publishers and law enforcement. In early 2012, it closed in the face of a lawsuit brought by 17 different publishers, but its digital archive was folded into that of LibGen, which became more popular as a result. This was an early example of what enforcement agencies now call “whack-a-mole” – the slippery, protean nature of shadow libraries makes them difficult to “burn down”.

It was around this time, in early 2010, that the precursor to Z-Library, called BookFi, was founded by two students at St Petersburg State University, Napolsky and Vladislav Chikhira. BookFi differed from its shadow-library peers in an important way. LibGen aimed to collect “academic literature” and to avoid “junky books” and “bestsellers”. Sci-Hub, founded by the Kazakhstani programmer Alexandra Elbakyan in 2011, likewise had a tighter focus on scientific papers, many of which had been produced using public money but were sold on by private companies with large profit margins.

BookFi, by contrast, had more generalist, Borgeesian ambitions. Its aim, Napolsky and Chikhira told a tiny Russian-language news website in 2011, was nothing less than to become “the largest online library” in the world, “encompassing fiction, scientific, reference and technical literature”. Z-Library began life as a simple search engine for ebooks, running off a student laptop on St Petersburg State University’s internal network. It quickly spiralled into an ambitious ideological project. Napolsky and Chikhira, both just 20 years old at the time, declared that “books are a public asset, and their access should be unrestricted”.

Today, Chikhira, judging by his social media accounts, lives a quiet life among a Russian exile crowd in İzmir, Turkey, where he moved after Russia invaded Ukraine in 2022. He works as a software engineer. By all appearances, he’s no longer involved in the Z-Library project, and he did not respond to my requests for an interview.

His fellow student founder, Napolsky, however, continued to work on Z-Library throughout the 2010s, as it increased in stature, expanding both its collection and its international user base. It was during this period that Valeriia Ermakova joined the project, and she and Napolsky started a romantic relationship. Information on the couple, including how and when exactly Ermakova came to the project, is scarce. (Repeated requests to interview the pair – through their US lawyers, former classmates and a source at Z-Library – were unreturned or rebuffed, citing the risk of influencing the ongoing case against them.) It is left to a florid piece of shadow-library fanfic, circulating

on social media, to imagine this moment: Ermakova, inspired by Z-Library’s mission, reaches out to Napolsky over the dark web in 2016, declaring that “I want to help you.” Soon, the couple are working together in the name of “a shared vision: THE LIBERATION OF KNOWLEDGE”.

As it expanded, moving into its second decade of operation in the early 2020s, Z-Library matured. It featured an easy-to-use search bar interface, surrounded by images of leather-bound volumes, projecting the comforting aura of an old-fashioned library. It had book recommendations, based on a user’s search history, and selections of books popular with other users, and even email alerts for users subscribed to particular topics.

Last year, an administrator of Z-Library, self-described as “one of the project’s oldest participants”, agreed to an interview on the condition of anonymity. (They would have met in person, but only in Russia or Belarus.) Over a number of months, I sent questions via email, and received detailed answers weeks later in .txt file attachments, described as “collective commentary from the Z-Library team”. They outlined to me what the mature phase of the organisation, still in operation today, looks like. “Roughly 20 people work on developing and maintaining the project,” they said, with a “team composition that is quite standard for IT products – development group, testing group, designer and technical support team”.

Though the anonymous administrator declined to tell me “absolute budget figures”, they claimed it would be “typical for an IT project of this scale”, all funded by small donations from their millions of users. Emails from Ermakova to a potential payments provider, accessed by the FBI and published in court documents, suggested that its monthly income in 2020 was around \$50,000, rising to \$150,000 during donation drives, which are mounted annually. The Z-Library administrator named the biggest expense as infrastructure maintenance – “3,000-5,000 processor cores and several petabytes of data storage” – followed by payroll for their team of “specialists” who keep the project running.

There are three main rules that such a project must abide by if it is to avoid enforcement action by western agencies, the founder of one Europe-based shadow library told me. The first is to operate within a niche, ideally one as specific as possible: no universalist ambitions, no mass-market books. Second, allow publishers and authors to make takedown requests, and comply with those requests quickly and comprehensively. Third, and perhaps most importantly, don’t allow money anywhere near the project. No donations, no tiered memberships, no subscriptions. As soon as money becomes a part of the system, the founder told me, you open yourself up to charges that you are profiting from stolen material – that you are not a library but a criminal enterprise.

By mid-2022, Z-Library was breaking all of these informal rules, by design, and it was beginning to attract attention. It solicited donations and offered a premium tier, for \$1, which allowed unlimited downloads of books (users were otherwise limited to 10 per day). Though the site nominally complied

with takedown requests, the crowdsourced nature of its archive, with hundreds of books uploaded by volunteers every day, made compliance difficult in practice. And, finally, because of its universalist aims, it was drawing in audiences beyond the millions of young students around the world who had traditionally used it to obtain textbooks and academic literature.

Among these new audiences were North American readers of young adult fiction, specifically fans of the phenomenally successful author Colleen Hoover, who had a new book out in 2022. That year, TikTok posts by Hoover fans promoted Z-Library as a remarkable new way to download her works for free. The posts clocked millions of views. Publishers' associations in the US and the UK began to devote more resources to investigating Z-Library.

Publishers' associations are, perhaps surprisingly, highly investigatory in nature and closely linked to police forces – the UK Publishers Association, for example, has a staff member on long-term secondment to the City of London Police Intellectual Property Crime Unit. According to a source familiar with the operation against Z-Library, researchers in one publishers' association assembled a dossier on the shadow library, documenting its domain names and, for the first time, definitively linking its activities to two named Russian individuals. These details were passed on to Interpol and the FBI. The source described a “stars-aligning moment”, with all the elements to entice US law enforcement agencies to take action to defend publishers' interests. There were US authors whose copyright had plainly been infringed, named individuals linked to the “international criminal networks” responsible and, finally, those individuals were set to travel outside of Russia in the near future, to a country with an extradition treaty with the US.

It seems that Napolsky and Ermakova had been linked to Z-Library due to a number of mistakes that look amateurish in hindsight. One of the ways the site took donations was via Amazon gift cards sent to Z-Library Gmail addresses. Napolsky had allegedly used his personal Gmail account as the back-up email for many of these donation accounts, and entered his name, phone number and address on Amazon accounts that cashed Z-Library donations. Likewise, Ermakova had registered her card details, and her St Petersburg address, with an Amazon account that redeemed donated gift vouchers, according to the complaint. All of these accounts were further linked by IP address data that showed that they usually logged in at the same time, from the same location.

Both Amazon and Google complied with an FBI warrant, providing information to investigators about Napolsky and Ermakova's account activities, and (in Google's case) the contents of their inboxes. For example, the Amazon account associated with Ermakova, which had cashed some Z-Library donations, had “placed more than 110 orders totalling over \$13,628.32, most of them for beauty and apparel products” over a three-year period, according to the complaint filed in 2022. It was presumably the Gmail inbox access that also alerted the FBI to Napolsky and Ermakova's travel plans. The night before they flew from St Petersburg to Córdoba, via Asunción, an application for arrest warrants against the pair was granted by the district court in New York.

“Z-Library has robbed individuals of the fruits of their labor,” wrote an FBI special agent in the warrant application, “for NAPOLSKY and ERMAKOVA's personal gain”. The next day, at 9.45am, the pair arrived in Argentina and the Interpol Red Notice was activated.

When Argentine police finally swooped on November 3, it was, according to the Z-Library administrator, “a complete surprise for all project participants”. Two weeks later, the pair were charged with copyright infringement, wire fraud, conspiracy to commit wire fraud, and money laundering conspiracy, and an extradition request was made. As the legal process rumbled on into the new year and throughout 2023, in both Argentine and US courts, the two Russians settled into house arrest in Córdoba.

By May 2024, the extradition process seemed to have stalled. Argentine courts had refused to approve it, waiting for more information from the FBI, and were also awaiting the result of an application by the Russians to the Argentine Supreme Court for refugee status. Meanwhile, courts in the US had been mulling over a memorandum filed the previous September by Napolsky and Ermakova's US lawyers, arguing that the charges themselves had been improperly filed by the FBI.

It was during this stalemate that Patronato de Liberados, an Argentine charity responsible for monitoring bail conditions, found the Russians missing from their assigned flat in Córdoba during a routine check. Another international arrest warrant was issued for Napolsky and Ermakova, and, in the absence of any trace of them passing through airports, the Spanish-language outlet *La Voz* speculated that they had gone underground in Argentina. However, in June last year, the administrator at Z-Library told me that they were both “currently safe in Russia”.

How had they managed to get there? No more details were forthcoming, either from Z-Library or their US lawyers, both unwilling to divulge anything that might affect the ongoing case. The source close to the investigation suggested that, in order for Napolsky and Ermakova to escape Argentina while they were subject to an international arrest warrant, the Russian state would have to be involved. “You can imagine something to do with Russia,” they said, “but I couldn't possibly comment on that.” One Spanish-language newspaper reported that Russian diplomatic staff were a constant presence at the Argentine court proceedings, even sometimes correcting court translators on points of legal vocabulary.

The left-humanist ethos of Z-Library doesn't seem to be a natural partner to Vladimir Putin's extreme rightwing, repressive dictatorship. Napolsky's co-founder Chikhira and many of their former classmates felt they had to leave Russia after its invasion of Ukraine in 2022. I asked the Z-Library administrator why they thought their project was allowed to operate safely under Putin. On that point, I received no answer.

A REVEALING ASPECT OF THE FIGHT BETWEEN shadow libraries and enforcement agencies is the obsession, on both sides, with terminology. The former tend to eschew talk of pirates and piracy. “For me, piracy means stealing to enrich yourself somehow,” the European shadow library founder told me. “I don't see our work in those terms.” But for those on the enforcement side,

any mention of the word “library” raises hackles. The preferred term, the source close to the operation to shut down Z-Library told me, was “organised criminal network”. These “criminal networks” were not to be considered in “cultural” terms, and had no “appropriate place in content ecosystems”.

These clashes over language are attempts to manoeuvre the battle on to favourable terrain. Enforcement agencies want this to be a simple case of theft, while the shadow librarians want a debate about ancient ideals and modern inequalities. How, for example, is it fair that an impoverished university student in China or India must pay thousands of dollars to highly profitable North American corporations simply to get online access to the textbooks or articles necessary for their education?

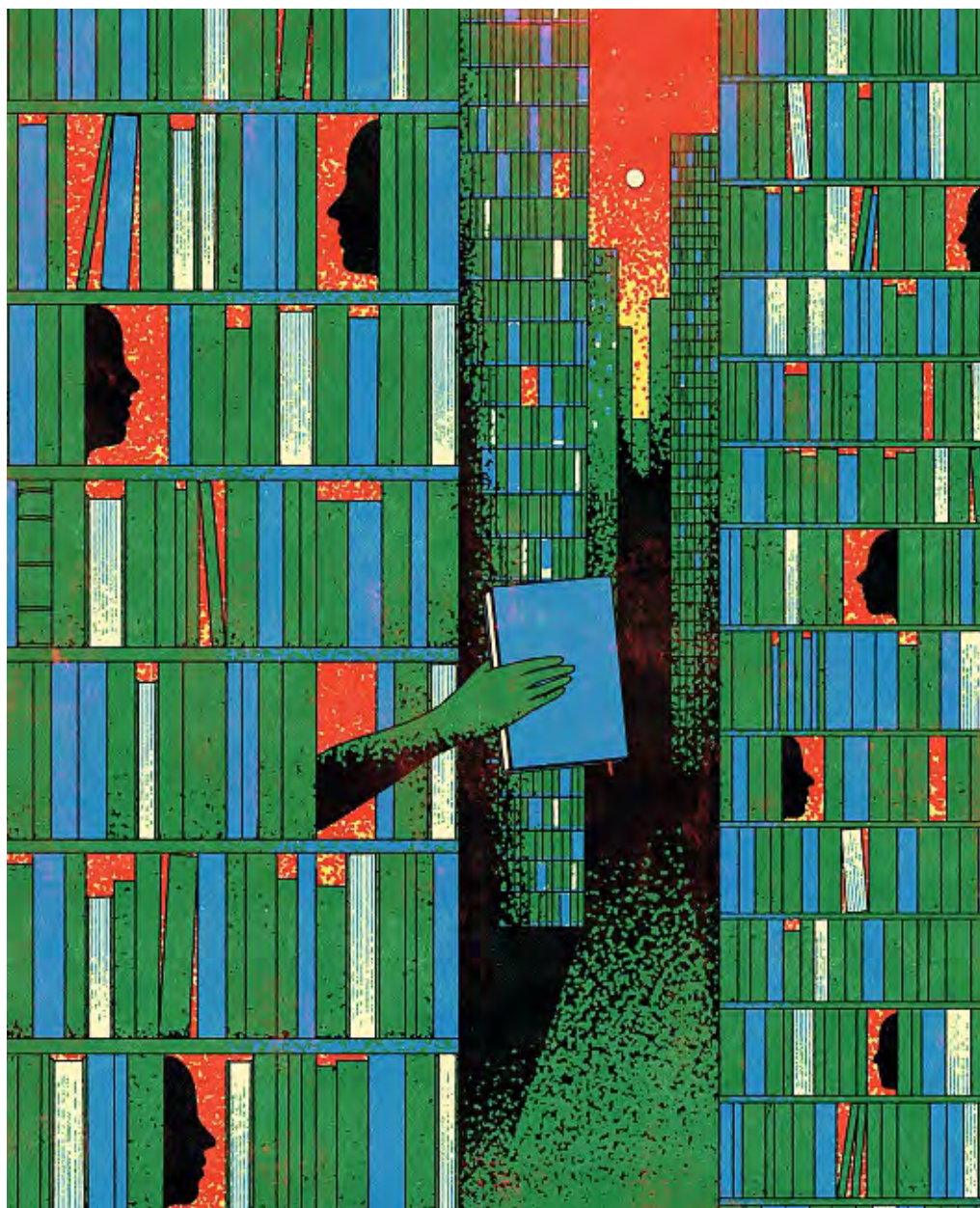
The takedown of Z-Library in 2022, though hailed as a triumph by publishers and law enforcement, revealed the extent to which shadow libraries had successfully set the terms of this debate.

Even authors who believed their earnings had been harmed by Z-Library drew a distinction between piracy of commercial fiction and that of academic literature, treating the former as harmful and the latter as almost sacrosanct. Hoover, whose pirated books may have initiated the whole operation, was moved to release a statement. “I was once a college student who couldn't afford textbooks,” she wrote in a Facebook post, which notably declined to condemn Z-Library. “I'd never willingly make that harder for people.”

What should be concerning for mainstream publishers is that, unlike the academic focus of Sci-Hub or LibGen, Z-Library wants to push the argument even further, away from the struggles of impoverished students and to the heart of the copyright contract itself. In the past this might have been easy ground for publishers, but in recent years it has become more fraught.

In the past two decades, around the world, author income has plummeted. In the UK, a series of depressing surveys by the Authors' Licensing and Collecting Society document this slide: from 2006 to 2022, they found that median earnings for authors had declined by just over 60 per cent. Meanwhile, the profits of the big publishing companies have increased. According to the ALCS, the publishing industry has become a microcosm of wider inequalities, “characterised by winner-takes-all dynamics and extremely high levels of earnings inequality”, with only a few percentage points of total publishing industry revenue making its way to authors at all. With median annual author income now at around £7,000, the question is increasingly who can afford to be an author, rather than who can write well enough or work hard enough to be one.

The argument that a global community of readers and writers should wholeheartedly support copyright enforcement because it funds literary production – because it enables authors from all backgrounds to write – is a compelling one. If, however, copyright is enforced with a militarised zeal akin to the war on drugs, and authors' earnings relentlessly squeezed, then the argument becomes a little harder to make. Z-Library's broad-brush claim that copyright is “a tool for profit extraction” that benefits “publishers not authors” will ring true for more and more people.



A suit earlier this year alleged that Google had actually used their Z-Library to train its Gemini model

The destruction of Z-Library in 2022 turned out to be merely an outage. After three months, its administrators announced its return, with a new system designed to evade domain shutdowns. This brief hiatus did, however, give rise to a new book piracy website called Anna's Archive, launched days after the Z-Library arrests, in a conscious response to them. Though it has many features of a library, this new organisation makes archival preservation its main aim: "to catalog all the books in existence"; ensure that they are never lost using duplicated, crowdsourced and geographically distributed storage; and to make them available to all. It has, by its own reckoning, collected about 16 per cent of all the books

and articles in the world, and 750,000 individual downloads are made through the site every day.

Anna's Archive might make publishers nostalgic for their old foe. Unlike Z-Library, with its documented history and messy paper trail, no one has been convincingly connected to it. Lawsuits against Anna's Archive go unanswered, and arrest warrants cannot be filed. And, while Z-Library's administrators agonise about moves towards "content commercialisation", unwilling to completely betray their communist roots, Anna's Archive is much more pragmatic. The organisation embraces the pirate identity, and anything that might help it attain its final goal of the total archive is considered: advertising, paid membership and high-speed paid access to its considerable archive for the training of corporate AI models.

A series of lawsuits in the past few years has revealed how pivotal shadow libraries have been to the training of LLMs. A suit earlier this year alleged that Google, once an informer on Napolsky and Ermakova, had actually used their shadow library to train its Gemini model. Anthropic, meanwhile, according to documents released as part of another suit, built its own internal shadow library out of stolen copyright material downloaded from pirate sites to improve Claude. (Anthropic agreed to settle with a group of authors and publishers late last year for \$1.5bn. Its general counsel Aparna Sridhar said the court's ruling meant that "training AI on books is fair use under copyright.") Suits against Meta surfaced documents showing that shadow library material had been used to train its Llama model, with authorisation from "MZ", which some speculate is Mark Zuckerberg. ("We respect third-party intellectual property rights," said a Meta spokesperson at the time, "and believe our use of information to train AI models is consistent with existing law.")

Anna's Archive, young and nimble, unconstrained by history, seems ready to capitalise on this opportunity. Documents from a recent lawsuit against the chipmaker Nvidia claimed that the company had been in talks with administrators from Anna's Archive to negotiate a fast download of their entire holdings, rather than laboriously crawling through captchas and limited download speeds. The pirate archivists are very open about the practice - blog posts often offer significant batches of written material to AI companies at a price, with conditions designed to give them a competitive edge, such as exclusive access to the material for a year following sale, before it is added to the public archive.

IN THE AMSTERDAM ZOO, BODÓ OUTLINED

two traditions of the internet to me, ones that he saw playing out in the AI gold rush today. On one side is the libertarian capitalist tradition, with its roots in Silicon Valley, which sees the entire corpus of the world's literature as a resource to be mined by machines to create profitable subscription services, automate jobs and concentrate wealth. On the other side is the anarcho-communist tradition, originating in grassroots mutual-aid projects in post-Soviet countries, which seeks to collect and preserve the world's literature for individual users to read, enjoy and learn from as they please. "Move fast and break things" versus "move fast and save things".

With the recent collaboration between Anna's Archive and Big Tech, these traditions seem, strangely, to have fused - each takes what they want from the deal, but what will the rest of us get?

The total library imagined by Borges almost 90 years ago was generative rather than acquisitive. It was not based on a successful effort to collect all books, but rather on the iterative combination of "all 25 orthographic symbols" until every possibility was exhausted and all books lay within its stacks. The result, he thought, would ultimately be disastrous: a "vast, contradictory Library, whose vertical wildernesses of books run the incessant risk of changing into others that affirm, deny, and confuse everything like a delirious god". **FT**

Felix Bazalgette is a writer and filmmaker from London