AI Is Eating the Book World

Jannis Lennartz

08.07.2026 doi: 10.59704/1f8f54bcf66a021f

In 2011, Marc Andreessen, a key figure in California's venture capital scene, coined the phrase: "Software is eating the world." The phrase describes the spread of software into everyday life and the displacement of physical business models. This process continues in an unexpectedly literal sense: AI companies purchase used books, scan them, and dispose of them to gather input for their models. Anthropic disclosed this practice earlier this year during a court proceeding. As the Washington Post reports: "The document describes how the scanning company's 'hydraulic powered cutting machine' would 'neatly cut' books, whose pages would later be 'scanned on high speed, high quality, production level scanners.' Finally, it notes, the scanning company will 'schedule with the recycling company to pick up the completed books.'" In Europe, reports are mounting regarding the purchasing practices of the Canadian company Zoom Books: The online bookstore is acquiring hundreds of thousands of nonfiction titles for which there is no longer a large market, but which represent previously unused material for training models.

The reason for this seemingly cumbersome method is the expectation that it will fall under the fair use provision of U.S. copyright law. It is a new stage of the AI copyright war: There is no flexible instrument such as the fair use doctrine in EU copyright law, but the opt-out structure of the exception and limitation for data mining might make used books fair game. Does one emphasize the creative or the destructive aspect of what Schumpeter called "creative destruction"? For some, book-devouring AI training signifies a capitalist fever dream: hungry machines exploiting the work of hundreds of thousands of creators and the cultural heritage of their readers, their billionaire overlords working towards a cyberpunk future known from books like Neuromancer or video games like Cyberpunk 2077. But concerns about the use or abuse of old books are also an indication of the problems facing European culture: The problem is not that Americans buy unwanted books, but that books are more and more unwanted in Europe.

Used Books as Fair Game?

Ownership of a physical copy of a book does not confer the right to reproduce it, neither in the U.S. nor in the EU. But the respective copyright laws use different instruments to balance the interests of rightholders and the public. In the US, this balance hinges on the fair use doctrine (Section 107 Copyright Act). Whether use without permission is fair depends on four factors, with a strong focus on the questions of how similar the use is to the work being used and whether they compete economically. In the EU, the right to use copyrighted material without the rights holder's consent depends on whether such use falls under one of the exceptions and limitations set forth in Article 5 Infosoc Directive. It gives preference to certain forms of use but does not allow for an exception on a case-by-case assessment beyond the written catalogue: a gain in legal certainty that comes at the cost of less fairness and flexibility in individual cases.

The debate over the legal limits of book digitization predates AI – it began with Google Books. Google scanned more than 40 million titles without the consent of the authors, although the company later implemented an opt-out clause. The company scanned works in the public domain – whose copyright protection had expired – as well as works that were still fully protected by copyright. Google shows previews of the books online, the corpus also functions as a basis for their Ngram Viewer. Google's action was adjudicated in Authors Guild, Inc. v. Google, Inc. and deemed to be fair use: "Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-

infringing fair uses. The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. Google's commercial nature and profit motivation do not justify denial of fair use." While things had quieted down around Google Books, AI companies' hunger for data has given the topic new significance. AI companies are searching for unused corpora and are well known for having resorted to reckless methods in their search for materials: for example, by using corpora from platforms with questionable copyright compliance, such as Libgen or Anna's Archive. Recent cases have been decided in favor of Meta and Anthropic. However, these decisions do not amount to a free pass, regardless of how the companies obtained the books. And, more importantly in practical terms: digital pastures have been grazed – physical books are their last frontier.

This raises the question of how the same practice fares under European law: Is the use of copyrighted works, whether digital or physical, illegal in Europe? The digitization of the book (as well as later copies in the learning process) constitutes a reproduction – one of copyright law's exploitation rights. It is only permitted if it falls under one of the exceptions and limitations of the Infosoc Directive. But there is no such exception or limitation. Alas, Art. 4 of the DSM Directive introduced an exception and limitation for text and data mining. It is now generally accepted that AI training falls under the definition of data mining. Its central feature is that it guarantees creators an opt-out. If they object to the use, the exception and limitation is no help. Anyone who opens a newly printed book will find an opt-out at the beginning. In the case of old books that the AI companies digitized themselves, neither the technological prerequisites for data mining nor the legal option to opt out existed. Therefore, an opt-out is not declared. This applies not only to truly old books, but also to those that are only a few years old but were printed before the DSM Directive introduced the exception and limitation for data mining. Are used books fair game? According to Recital 18, it is possible to declare the opt-out not in the copy but by other means. This gives publishers the ability to prevent even books that have already been published before the DSM Directive from being used for data mining. But some publishers have gone out of business and some authors have passed away. Not every author has a legal department to look after them. Without an opt-out, it all depends on the unclear meaning of the work being "lawfully accessible" (Art. 4 (1) DSM Directive). A wide interpretation is possible on the level of the DSM Directive and its national implementation in Germany. But its compatibility with European and national fundamental rights does not seem self-evident: after all, this involves a significant encroachment on authors' rights.

The Whimpering of Hollow Men

In America, old books are midwifing the machine god. What about Europe? Lamenting the use of old books as abuse feels like compensatory behavior: Europeans begrudge American machines the books in which they themselves have long since lost interest. We are witnessing an unmitigated crisis of literary culture in publishing, education and politics.

A few examples illustrate the changes that have taken place since the middle of the last century: after Thomas Mann wrote Doctor Faustus while in exile in California, he went on to write a commentary on its creation – a miniature of the literary culture of modernity in which art was *high* art: difficult, demanding, grand. Today, love stories between werewolves and vampires popularized on TikTok, or cross-selling fantasies like Save Me/Maxton Hall, may generate revenue for publishers and some self-publishers in the young adult category. But only from a philistine's perspective does reading have value as such. The continued existence of books as a medium masks the cultural rupture. A rupture aided and abetted by the state. Two examples from Germany: students' reading proficiency continues to decline; according to the IGLU study, one-quarter of elementary school students do not meet the minimum standard. At the same time, classics are being distorted in Berlin schools by being read in so-called "simple language." This decline in people's intellectual capacity is reflected by their

representatives: during a visit to Colombey-les-Deux-Églises, Konrad Adenauer asked President de Gaulle if he could spend an hour alone in his library – to get to know the French politician. Today, former chancellors like Angela Merkel or Olaf Scholz are considered well-educated if, while traveling abroad, they tell journalists that they are reading a book on the history of the country they are visiting. Meanwhile, Thuringia's Minister President, Mario Voigt, is reported to have his speeches partly generated by AI – even on an issue like the Holocaust. Why bother? He leaves the reading of books to American machines.

ABOUT THE AUTHOR

Jannis Lennartz
Jannis Lennartz is a Privatdozent at Humboldt University in Berlin.

SUGGESTED CITATION

Lennartz, Jannis: AI Is Eating the Book World, Verfassungsblog, 08.07.2026, doi: 10.59704/1f8f54bcf66a021f.

Licensed under CC BY-SA 4.0.

Langzeitarchiviert bei IntRechtDok unter: doi.org/10.17176/20260708-190519-0.