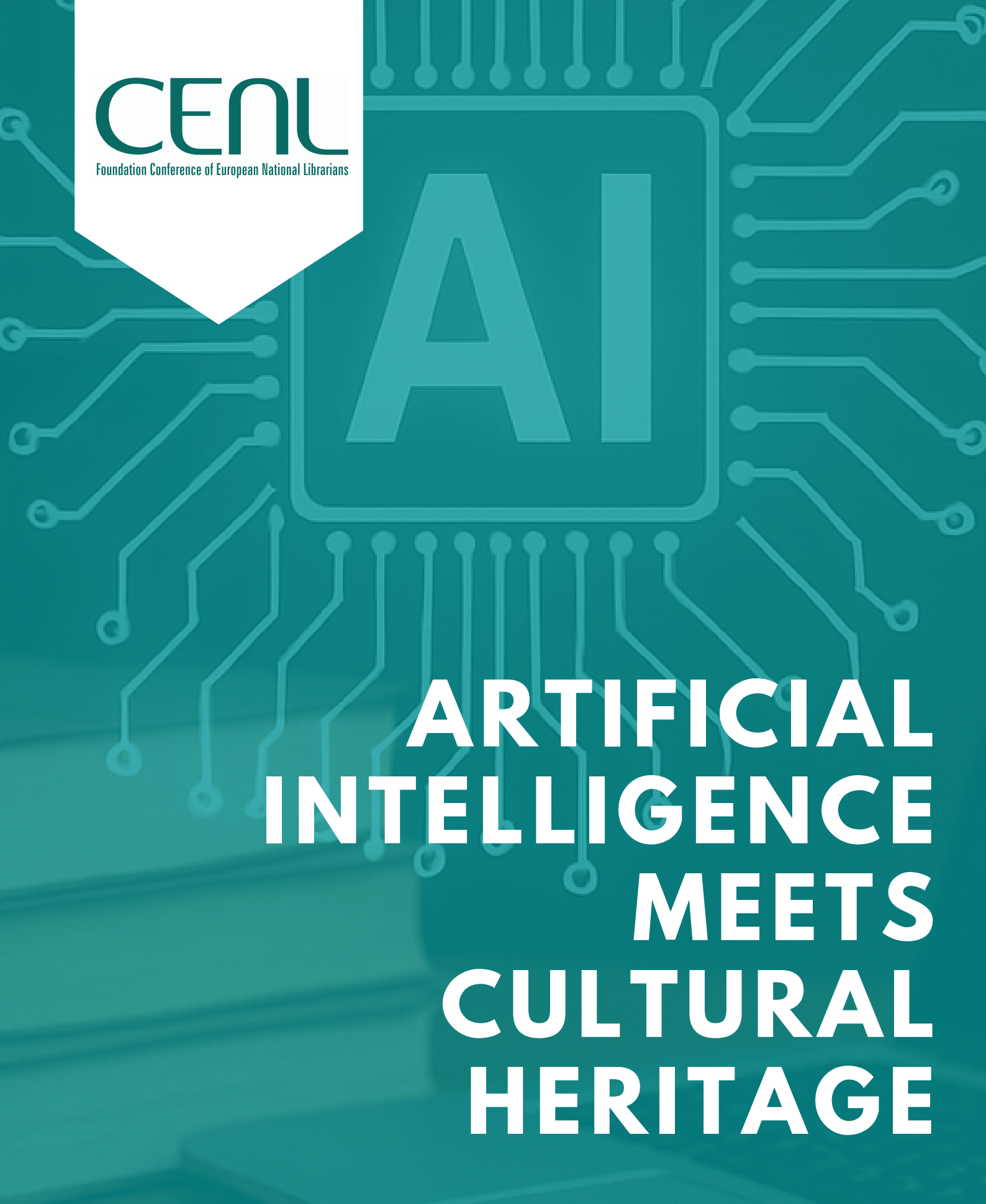
CENL
Foundation Conference of European National Librarians



ARTIFICIAL INTELLIGENCE MEETS CULTURAL HERITAGE

WWW.CENL.ORG

CENL
Foundation Conference of European National Librarians

Artificial Intelligence Meets Cultural Heritage

The Transformative Power of AI for and through National Libraries

Artificial Intelligence Meets Cultural Heritage -
The Transformative Power of AI for and through National Libraries
Conference of European National Librarians (CENL)
edited by Susanne Oehlschläger and Annika Wenzel, 2025

CENL Secretariat
c/o Deutsche Nationalbibliothek (Leipzig, Frankfurt am Main)
Deutsche Nationalbibliothek, Adickesallee 1, 60322 Frankfurt am Main, Germany
www.cenl.org

<urn:nbn:de:101-20250321144>

Content

Preface

AI is everywhere. This is not new - but the rise of ChatGPT in late 2022 has undoubtedly made AI one of the most widely discussed topics worldwide. What is new is the unprecedented pace of development, with new models emerging almost daily. Unsurprisingly, this technology has also reached the doorsteps of national libraries.

As leaders of Europe's national libraries, we are confronted with many shared questions: How can we make cultural heritage accessible to wider and more diverse audiences? How can we ensure the long-term availability of our digital collections? And how can we do so sustainably—both environmentally and institutionally? In the context of AI, further questions emerge: What is possible? In which areas can AI support our work? Where can it meaningfully enhance our services? And conversely, what should national libraries not do with AI?

CENL's Artificial Intelligence in Libraries Network Group has been highly active for six years, and its continuation for another three years reflects the shared commitment within our community. Yet it was clear that the directors of Europe's national libraries wished to engage even more deeply with the opportunities and challenges of AI. For this reason, the 2025 Annual General Meeting (AGM) of the Conference of European National Librarians (CENL), hosted by the National Library of Scotland in Edinburgh, was dedicated to AI.

The AGM's motto, "Artificial Intelligence meets Cultural Heritage," serves not only as a title but also as a statement of intent: national libraries, as stewards of invaluable cultural heritage, are both willing and equipped to explore the meaningful application of AI. Many libraries do not see AI as a threat but as a tool—one that enables us to carry out essential, automatable tasks more quickly and efficiently. During the session "AI in Action: Voices from National Libraries," several CENL members shared their hands-on experiences, inspiring others and offering ideas for implementation within their own institutions.

As in previous years, I am pleased to announce that CENL is publishing most of the papers that contributed to our fruitful discussions. While these papers cannot fully replace the rich debates of our AGM, they provide valuable insights into the current thinking and experimentation across our community.

On behalf of the CENL Executive Committee, I extend my sincere thanks to all contributors and to our hosts in Edinburgh. Thank you for your engagement, your ideas, and your hospitality.

Frank Scholze
Director General of the German National Library
Chair of CENL

Melissa Terras:
Be More Library: Upholding Library Values in a Tech Industry World

In an era of rapid technological advancements, and with the recent hype and disruption that Generative Artificial Intelligence (GenAI) has brought to our information environment, libraries today find themselves wrong-footed. Our enduring memory institutions, known for their foundational values of accessibility, inclusivity, and preservation of knowledge, are increasingly challenged by a technology industry driven predominantly by economic gains and extractive capitalist approaches that deprioritise trustworthy information practices. Library staff, and library boards, are repeatedly asked to immediately respond to recent advancements, opportunities, and threats posed by AI. However, they rarely have the education, resources, or capabilities to keep up to date in this fast-moving commercial area, or to think through the ramifications of embedding particular AI tools and systems within library data flows and processes. There is a sense of panic in the sector as we now routinely see openly licensed collections, put online to benefit human readers and increase access, be scraped and repurposed by bots to feed the AI machine, often crippling our online services[1] but without giving credit nor recompense to the institutions who provide them, which somehow feels like an affront to all we stand for. However, in this short essay, based on my keynote for the 2025 Conference of European National Librarians (CENL) conference in Edinburgh in June 2025[2], I argue that the pressing question is not what libraries should do about AI, but how libraries can reaffirm their core principles and assert their essential role in safeguarding the information environment itself. The answer is not to react immediately and be trapped in perpetual responsive mode to AI, but to enact guiding principles that have underpinned library activities for over a century, to allow us to flourish, and confidently prioritise the activities our organisations undertake.

Understanding the Threat

At the heart of this AI moment is a clash of values. The field of Valuation Studies tells us that there are three main types of value: Economic (financial but also political value); Social (activities that are "good" and valuable to human life and societal flourishing); and Semiotic (things that signify values, see consumer culture theory on branding and meaning)[3]. Traditionally, libraries have focused on social values such as cultural preservation, access, community service, and collaboration. I delight in the fact that libraries are one of the few places in modern society where it is socially acceptable to visit and spend no money at all.

[1] Weinberg, M., June 2025. "Are AI Bots Knocking Cultural Heritage Offline"? Glam-e Lab. Engelberg Centre on Innovation Law & Policy. NYU School of Law https://www.glamelab.org/products/are-ai-bots-knocking-cultural-heritage-offline/, Last visited: 5 December 2025

[2] https://www.cenl.org/annual-general-meeting-2025/, Last visited: 5 December 2025

[3] Karababa, E. and Kjeldgaard, D., 2014. Value in marketing: Toward sociocultural perspectives. *Marketing Theory*, *14*(1), pp.119-127 https://doi.org/10.1177/1470593113500385, Last visited: 5 December 2025

In contrast, the technology industry's values, originating from the disruptive ethos of the 1980s[4], often clash with those of libraries. Principles like unlimited access to computers and compute, and mistrust of authority, pose significant threats to established institutions. Technology companies prioritise decentralisation and disruption, often targeting the very authorities — such as libraries — that they perceive as traditional and static. The very business of libraries, who champion social good, while providing trusted environments to store information about human society on a much longer-term, stands in direct opposition to current tech industry priorities and values, such as Facebook's original motto "move fast and break things"[5]. Our modern technology industries prioritise economic value and its associated political power, and their activities increasingly intertwine the two, prioritising disruption and disinformation as casual side effects in the pursuit of financial profits and political power[6]. In a matter of two years, a few tools that provide the latest technology product, GenAI, have casually and irrevocably polluted the online information environment that took humans three decades to construct worldwide[7], while encouraging the rise of the far-right, and threatening democracies worldwide[8]. This clash of values that libraries and librarians are now contending with is not only theoretical.

Defining and Upholding Library Values

But what are Library values? To understand this, in advance of the CENL 2025 conference I accessed an attendee list, which showed 38 libraries throughout Europe would be present. I then tracked down their mission statements, which in some cases was simple (shout out to the National Library of Scotland who provides theirs in both PDF and in text format[9]) but in others involved digging through papers or legal statutes to understand the institution's current guiding principles. My first Library Recommendation is: *make it easy for your community to access a statement that sets out your core values*.

Once these statements had been located, AI assisted translation was used to convert all into English, giving a corpus of 150,000 words. Voyant Tools[10] allowed a simple exploration of the texts and their messaging. There were 8991 unique words, with the most frequent words in the corpus being: library

---

[4] Levy, S., 1984. *Hackers: Heroes of the computer revolution*. Garden City, NY: Anchor Press/Doubleday

[5] Blodget, H. (October 1, 2009). "Mark Zuckerberg On Innovation". *Business Insider*. https://www.businessinsider.com/mark-zuckerberg-innovation-2009-10?r=US&IR=T, Last visited: 5 December 2025

[6] Suarez-Villa, L., 2016. Globalization and technocapitalism: The political economy of corporate power and technological domination. Abingdon: Routledge

[7] Adami, M. (26th November 2024). "AI-generated slop is quietly conquering the internet. Is it a threat to journalism or a problem that will fix itself?" Reuters Institute, University of Oxford https://reutersinstitute.politics.ox.ac.uk/news/ai-generated-slop-quietly-conquering-internet-it-threat-journalism-or-problem-will-fix-itself, Last visited: 5 December 2025

[8] "the crisis in information which underpins the crisis in democracy which are all a result of the revolution in communication brought about by new technology platforms. All of which is fuelling the far-right populist politics that are spreading across the world." Cadwalladr, C. (May 12, 2025). "A secret US-UK sweetheart tech deal?" *How to Survive the Broligarchy.* https://broligarchy.substack.com/p/a-secret-us-uk-sweetheart-tech-deal, Last visited: 5 December 2025

[9] https://www.nls.uk/about-us/who-and-what/strategy/, Last visited: 5 December 2025

[10] https://voyant-tools.org, Last visited: 5 December 2025

(2132); national (1338); libraries (806); digital (753); development (722). This shows the importance of digital to the modern library environment which should be of no surprise to anyone working within the sector but is often a surprise to those who do not. In amongst these 150,000 words, AI was only mentioned 15 times across 10 documents: a strategic response to AI is not embedded into most libraries' strategy, despite digital being foregrounded. The average words per sentence in these statements was 31.4, which indicates long, hard to follow constructions: instead, libraries should *find a way to express what values underpin your institutional work simply and succinctly*.

A thematic analysis of the 300 most common words in the corpus detailed three main institutional focuses (the number in brackets shows how many times the word appears in the corpus):

Knowledge and Learning:
High Frequency: "library" (2132), "information" (543), "research" (456)
Related Concepts: "education" (269), "training" (275), "reading" (304)

Culture and Heritage:
High Frequency: "cultural" (597), "heritage" (414), "culture" (332)
Related Concepts: "preservation" (115), "tradition" (implied in similar terms like "heritage")

Accessibility and Inclusion:
High Frequency: "access" (397), "public" (502), "accessible" (153)
Related Concepts: "users" (281), "inclusive" (implied through accessibility terms)

Using Topic Modelling (a process used in natural language processing to identify themes within a set of documents) to distil the values contained within the documents gives a further common roadmap across the 38 institutions of what national libraries prioritise:

- Enhance Access and Inclusivity
- Strengthen Cultural Preservation
- Promote Knowledge and Learning
- Drive Digital Transformation
- Foster Community and Collaboration
- Ensure Sustainable Management
- Enhance Strategic Planning
- Excellence

Most of this list details social values – things that improve and support human society – with the last "excellence", being semiotic: in National Library statements everything they do must signal excellence and uphold certain standards.

Moving Forward in an AI World with our Values at Heart

How can we move forward in an increasingly AI world while upholding library values?

Firstly, we must understand that AI should not be viewed as an isolated issue but as an integral and pervasive challenge affecting all sectors, including libraries: all online data is being integrated in ways we may not have intended and cannot untangle. In one small section of Kate Crawford and Vladen Joder's 2023 work "Calculating Empires, a Genealogy of Technology and Power Since 1500"[11] the interplay between library digitisation and AI is mapped, illustrating how library resources contribute to AI datasets, particularly in natural language processing and computer vision, with profound implications extending even to military applications. The cat is out the bag. In response, libraries must be ready to critically evaluate AI's impact and enter continuous dialogue about its ethical use and the potential misuse of data by malevolent actors. We must interrogate our attitude to open data, repurposing the FAIR data practices of "as open as possible, as closed as necessary" that were designed to protect individuals[12] so that we also protect institutional resources. We can do so while exploring sustainable business models and potentially developing our own proprietary AI tools, developing shared goals, and leveraging collective expertise to foster innovation without compromising core values.

We must realise that *we cannot rely on governmental support to uphold our library sector values*. Many politicians and civil servants, particularly in Western democracies, have made it clear that they are leaning into the AI hype promise of destabilisation and economic growth, to create value for everyone – as long as it is economic value[13]. They have stressed that social aspects of technology are not a priority[14]. That means it is up to librarians to understand the long-term implications of governmental short-term decisions about AI, and we need to understand what the impact of these tools are on information environments, and to safeguard cultural heritage content because, quite simply, no-one else is going to.

Part of the disruption of the tech companies is an attack on copyright, which is an inconvenience to GenAI, given it requires access to ever more modern, in copyright, datasets. There were reasons why Dr Carla Hayden was sacked from the Library of Congress in 2025, which was not only about just diversity and inclusion. She and Shira Perlmutter, the former Register of Copyrights and Director of the U.S. Copyright Office, were standing in the way of the technology companies getting hold of

---

[11] https://calculatingempires.net, Last visited: 5 December 2025

[12] Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L.O.B. and Roos, M., 2020. The "A" of FAIR–as open as possible, as closed as necessary. *Data Intelligence*, 2(1-2), pp.47-55. https://direct.mit.edu/dint/article/2/1-2/47/9998/The-A-of-FAIR-As-Open-as-Possible-as-Closed-as, Last visited: 5 December 2025

[13] Booth, R., 12th January 2025. "'Mainlined into UK's veins': Labour announces huge public rollout of AI Plans to make UK world leader in AI sector include opening access to NHS and other public data" Politics, *The Guardian*. https://www.theguardian.com/politics/2025/jan/12/mainlined-into-uks-veins-labour-announces-huge-public-rollout-of-ai, Last visited: 5 December 2025

[14] Amnesty International, 10th July 2025. "UK: Government's unchecked use of tech and AI systems leading to exclusion of people with disabilities and other marginalized groups". https://www.amnesty.org/en/latest/news/2025/07/uk-governments-unchecked-use-of-tech-and-ai-systems-leading-to-exclusion-of-people-with-disabilities-and-other-marginalized-groups/, Last visited: 5 December 2025

digital contents to feed the AI machine[15]. This is relevant to the library sector as more than half of CENL libraries are legal deposit libraries, whose very existence is tied to copyright frameworks: The GenAI attack on copyright is an attack on libraries. Where are we talking about the value of copyright? Where are libraries' public statements standing up not only for creatives[16] but also the means by which we can collect human history? If it is not possible to speak out externally because of very real potential ramifications, what are institutions doing internally, or between trusted colleagues, to withstand these attacks and understand the rhetoric, and the very real threats, our copyright institutions are up against? Additionally, how do we make sure our boards, managers, and staff *are kept up to speed in this rapidly changing space*?[17] Can we resist overtures from AI companies offering us money for privileged access to our collections[18], at a time of economic austerity when we could do with the extra funding given our governments don't care about our social values? Can we collectively call out "library-washing"[19] and understand what we may lose when we align heritage collections with GenAI activities, and what the technology companies may gain? To counteract this, can *libraries start to act as a network to share information, training, examples, and build best practice, in a way that cooperatively upholds our own values*?

The good news is that the business model of GenAI is looking more and more fragile, with a sense Summer 2025 sees as at the height of a market bubble[20]. A resettling of the AI space will hopefully stop prioritising GenAI over other applications of AI, and it should be stressed that AI is not bad within itself: *there are so many shared data problems in the library sector that would benefit from AI*, for example, cleaning up metadata, the creation of linked open data, identifying bias in collections, improving poor OCR and HTR, etc[21]. The library sector needs to come together and identify where it could most usefully apply AI, in an ethical manner, on our own content, and what it would take to build our own tools and integrate it with our own systems. We also need to, very rapidly, find a way to document and preserve the effect AI is having on the information environment, otherwise in only a few years it will be impossible to understand the changes AI made to our world

---

[15] Nawotka, E. 12th May 2025. "Copyright Chief Fired Amid AI Debate". *Publishers Weekly*, https://www.publishersweekly.com/pw/by-topic/digital/copyright/article/97748-copyright-chief-fired-amid-ai-debate.html, Last visited: 5 December 2025

[16] https://newsmediauk.org/make-it-fair/, Last visited: 5 December 2025

[17] Agnese, P., Arduino, F.R. and Di Prisco, D., 2025. "The era of artificial intelligence: what implications for the board of directors?". *Corporate Governance: The International Journal of Business in Society*, 25(2), pp.272-287

[18] Thomson, E. 5th March 2025. "OpenAI expands AI reach with $50M NextGenAI partnership featuring Oxford, Harvard, and MIT". https://www.edtechinnovationhub.com/news/openai-expands-ai-reach-with-50m-nextgenai-partnership-featuring-oxford-harvard-and-mit., Last visited: 5 December 2025

[19] Coined by Professor Paul Gooding at CENL 2025. AI companies have been known to take the rhetoric of the open movement to advertise their activities, without upholding open-science principles. Are they now doing the same to libraries? See Liesenfeld, A. and Dingemanse, M., 2024, June. "Rethinking open source generative AI: open-washing and the EU AI Act". In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 1774-1787) https://doi.org/10.1145/3630106.3659005, Last visited: 5 December 2025

[20] Thornhill, J. August 21st 2025. "Brace for a crash before the golden age of AI" Opinion, Artificial Intelligence. *Financial Times* https://www.ft.com/content/a76f238d-5543-4c01-9419-52aaf352dc23, Last visited: 5 December 2025

[21] Jaillant, L., Warwick, C., Gooding, P., Aske, K., Layne-Worthey, G. and Downie, J.S., 2025. *Navigating artificial intelligence for cultural heritage organisations*. London: UCL Press https://doi.org/10.14324/111.9781800088375, Last visited: 5 December 2025

in the early to mid-2020s[22]. Libraries and librarians can do all this ourselves by building a community and having a set of shared goals about AI. It is up to Libraries and Librarians to do this, upholding our own values as we understand, document, and roll out AI. No-one else will uphold our social values for us in the AI space.

Case Study: Transkribus

The call for community in building our own AI tools is not theoretical. For the past eight years, Transkribus[23] has provided a range of machine-learning based Automated Text Recognition tools to those in the library and archive sector. Originally released as part of an European Union funded project[24], at the end of its funded phase in 2019 Transkribus legally instantiated as a cooperative, READ-COOP[25], which is now co-owned by over 250 different individuals, libraries, and archives worldwide. All profits are reinvested in the infrastructure, a financial structure which also allows it offer free access to early career researchers. The winner of the Horizon Europe Award for Impact in 2020, Transkribus has 350,000 registered users, who have now accurately transcribed 100m images of historical documents, which have underpinned hundreds of research projects, scholarly editions, and academic outputs[26]. This cooperative approach highlights the potential of shared models to create sustainable, socially oriented AI tools, for a particular focus, community, and audience[27]. This stands in stark contrast to the profit-first, winner-takes-all mentality pervasive in the wider technology sector. It also offers a tangible example of how libraries can wield technology to support their values, ensuring that the community works together to serve collective rather than individual goals.

Be More Library: A Roadmap for AI Library Action

When discussing what to say in my CENL keynote with Dr Sarah Ames, the Digital Scholarship Librarian at the National Library of Scotland, Sarah remarked that she didn't understand why the sector was so panicked about AI, given every few years there was always another disruptive

[22] Padilla, T. 4th April 2025. "Preserving AI: Future scholars will need an enduring digital record of AI technology" *Inside Higher Ed*, https://www.insidehighered.com/opinion/views/2025/04/04/ai-preservation-unaddressed-challenge-opinion, Last visited: 5 December 2025

[23] https://www.transkribus.org, Last visited: 5 December 2025

[24] Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S. and Gatos, B., 2019. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of documentation*, 75(5), pp.954-976 https://doi.org/10.1108/JD-07-2018-0114, Last visited: 5 December 2025

[25] https://readcoop.org Date last visited: 26th August 2025

[26] Nockels, J., Gooding, P., Ames, S. and Terras, M., 2022. Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival Science*, 22(3), pp.367-392 https://doi.org/10.1007/s10502-022-09397-0, Last visited: 5 December 2025

[27] Terras, M., Anzinger, B., Gooding, P., Mühlberger, G., Nockels, J., Romein, C.A., Stauder, A. and Stauder, F., 2025. The artificial intelligence cooperative: READ-COOP, Transkribus, and the benefits of shared community infrastructure for automated text recognition. *Open Research Europe* https://open-research-europe.ec.europa.eu/articles/5-16, Last visited: 5 December 2025

technology coming down the pipes: the answer was always to "Be More Library" about it. This led me to consider what would "being more library" mean, in relation to AI?

Firstly, *libraries should* ***reaffirm their social values***, maintaining a steadfast focus on social values over economic ones, ensuring library initiatives are informed by a commitment to societal benefit rather than short-term financial gains. Secondly, going back to the library value that centres **community and collaboration:** *libraries should build coalitions across institutions to develop shared goals, focusing on AI's ethical use and leveraging collective expertise to foster innovation without compromising core values*. By emphasising their collective roles as data stewards, libraries can ensure their content's integrity and longevity. Thirdly, libraries should ***advance digital literacy around AI***: equipping library staff and users with the skills to navigate and critically evaluate AI technologies, emphasising transparency and responsible use. Fourthly, even though libraries are used to being at arms' length from political processes, they should start to ***engage in policy reform***, advocating for copyright reforms that reflect digital realities while protecting cultural heritage, pushing back against legislation that endangers their core mission. Finally, they should *promote ethical AI use*, seeing AI as a tool just like any other, and developing and prioritising approaches that place community benefit over financial or political profit, while weighing up the environmental and social costs that AI levies.

Conclusion: Be More Library

Faced with the pressures of a world increasingly dominated by technological forces, libraries must lean into communicating their foundational values and respond to AI with those in mind. By prioritising inclusivity, cultural preservation, and community collaboration, libraries can navigate the recent challenges posted by GenAI with integrity. The call to "Be More Library" asks us to uphold the principles that define our memory institutions, ensuring they remain indispensable pillars of knowledge and democracy in a rapidly changing world. It is up to the library community themselves to project the semiotic values of trust, longevity, and indeed excellence, that characterise our aims as well as our activities. Asserting the unique role of libraries will allow them to commit to social values that have long defined them, even as they embrace ongoing opportunities and challenges of the digital age.

Paul Gooding:
AI in the National Library: Thinking Fast and Slow

Introduction

Good afternoon, everybody, and I am honoured to join you for your annual gathering. Many thanks to Amina Shah and the National Library of Scotland for inviting me to speak. In this talk, I will be talking about how the library sector can ground itself in response to the hype around new Artificial Intelligence technologies, through the metaphor of "Fast and Slow AI." A lot of what I will discuss comes from a decade and a half of working with the UK Legal Deposit Libraries, and so is a little UK-biased. For that I apologise. This talk will elaborate on the role of Fast and Slow AI as a response to the need to centre users and libraries in our adoption of AI, bringing together a mixture of theory that has inspired my thinking, and findings from my own research.

Marshall McLuhan (1967) used the metaphor of the "figure" and the "ground" to explain how society can often miss the wider significance of a technology. The figure is that which leaps out at us, and grabs our attention, while the ground is the context that shapes and influences that figure. To understand, any technology, then, we need to examine the ground or context within which it exists. I will address each in turn: first, the figure of Artificial Intelligence, broadly understood, as it is applied to libraries; and second, the "ground" of data capitalism, and how the tensions between extractive AI practices and the values and work of libraries might inform our response to the technologies.

The Figure: Artificial Intelligence in Libraries

Artificial Intelligence can be defined in a variety of ways, but the UK government defines it in simple terms: "machines that perform tasks requiring human intelligence, especially when machines learn from data how to do these tasks" (Office for Artificial Intelligence, 2021). This definition cuts through a great deal of the hype around "AI", emphasises it as an umbrella term for a range of interrelated and overlapping technologies, such as machine learning, deep learning, natural language processing, expert systems, robotics, and computer vision. None of these things are new; indeed, when individuals such as Linda C. Smith (1976) established AI as a research area in Information Science in the 1970s, they were talking about similar AI applications.

In the same way that AI in libraries is not new, nor are debates around technological adoption in libraries. Library catalogues (Coyle, 2017), CD-ROMs (Lambert and Ropiequet, 1986; Bevan, 1994), and more recently the Blockchain (Bashir and Warraich, 2023), have all led to debates about how we leverage new technologies in a way that works for all of our communities. However, when exposed to what TD Wilson (1981) refers to as the "user's life world" – the "totality of experiences centred upon the individual as an information user" – the implementation of new technologies in libraries often fall short of the initial hyperbole that surrounds their introduction. Chern Li Liew's (2012) words

from 2012 still ring true to me: "one gets the feeling that the current system with digital initiatives is that a *big digital information party* is being planned but somehow, in the excitement, someone is being left out." In my own research, I've been struck by the way in which certain user communities are often marginalised or poorly served by technologies that were introduced to enhance access and usage of our collections.

From 2010 to 2016, I set out to understand how digitised newspapers were being used by various user communities, using the digitised newspaper collections of the National Library of Wales[1] and The British Library.[2] However, when I surveyed users I was surprised to find stories of exclusion alongside those of innovation and promise, caused by commercial digitisation agreements that placed materials behind paywalls. This most severely affected academics in precarious employment, who lost institutional access when their contracts expired, less well-resourced universities that couldn't afford to provide access, and members of the public in rural and deprived areas of England whose public libraries were unable to maintain subscriptions (Gooding, 2017). These findings remind us of what Manuel Castells (2002, p. 262) warned in 2002: "Key urban centres, globalised activities, and the higher educated social groups are being included in the internet-based global networks, while most people and regions are switched off."

Similarly, from 2017 to 2019, I led a first-of-its-kind project to understand the impact and usage of Non-Print Legal Deposit (NLPD) (*The Legal Deposit Libraries (Non-Print Works) Regulations 2013*, 2013). Again, it emerged that while the introduction of electronic legal deposit was met with great enthusiasm, many end users felt dissatisfied with their experience. Library staff were enthusiastic about the value of the NPLD collections but reported that usage seemed lower than they would expect. As we investigated further, we found very little evidence of user involvement in the development of the UK NPLD regulations, or the subsequent implementation of the service. As one library interviewee argued:

> *When we first started thinking about implementation, we were so consumed with the collection of the material… at no point is the end user considered. The end user is never considered in the regulations so it's never been at the forefront. And all of a sudden we start implementing and realise that the user is not considered at all (Gooding, Terras and Berube, 2019).*

When we talk to our communities, we often find that our technologies, and by extension our services, aren't working for them in the real world. They are being left behind, because we sometimes license materials in ways that limit access, because we don't always talk to them until it's too late, and because we often forget to give them agency to decide what they want or need. This led me to last year where I led a project entitled "Inclusive Requirements Elicitation for AI in Libraries." Funded by the Arts and Humanities Research Council under the Bridging Responsible AI Divides scheme, it aimed to develop a model for responsible AI systems development in libraries seeking to include

[1] Welsh Newspapers Online: https://newspapers.library.wales/, Last visited: 5 December 2025

[2] British Library Nineteenth Century Newspapers, now part of Gale Primary Sources: British Library Newspapers: https://www.gale.com/primary-sources/british-library-newspapers, Last visited: 5 December 2025

knowledge from Indigenous communities. The team contained a mixture of Indigenous and non-Indigenous researchers, including partners from Information Studies at the University of Glasgow, King's Digital Lab at King's College London, the Jumbunna Institute for Indigenous Research and Education at University of Technology Sydney, and the Digital Preservation Coalition.

Requirements elicitation is a widely used term in business analysis for identifying and prioritising requirements for a business change. It involves the elicitation of functional and non-functional requirements for the creation of a new system or process, generally through a process of engagement with key stakeholders that have a relationship to the proposed new system. Our intention was to intervene in how requirements elicitation is undertaken in systems utilising Indigenous knowledge, understanding that Indigenous peoples are a key stakeholder in such systems (Bouich *et al.*, 2025). Indigenous knowledges are held by 476 million Indigenous peoples across 5,000 distinct communities worldwide, and comprise differing systems of how each community approaches knowledge; how it is produced, how it is disseminated, and how it fits into that community's wider worldview. These forms of mediation, and their significance within a particular community, are not global, and can only be understood through a process of developing cultural literacies in relation to specific communities. Often such ways of understanding are relational. As Dudgeon (2023) argues, relationality means centring a worldview that stresses a harmonious relationship between people and the planet, and thus requires us to understand human relationships to knowledge and technology.

The Indigenous Archives Collective Position Statement on the Right to Reply to Indigenous Knowledges and Information Held in Archives (2021), emphasises the community's right to know, and the closely linked right to reply, about the reuse of community knowledges, the need to ensure participation and consent from communities, and the need to create culturally safe environments for all participants. To understand how these rights and needs related to the design of AI systems within libraries, we undertook three workshops. Workshop one immersed attendees in this relational framework by introducing participants to Indigenous data governance principles and the application of AI in libraries, before attempting to map out key decision-making points for Indigenous community input into AI processes. Workshops two and three built on this via an action research approach, by developing an LLM dashboard ('kingsdigitallab/iREAL', 2025) applied to a specific dataset of school records from Western Australia – many of which held descriptions of the treatment of Indigenous pupils. At workshop two, the LLM was explored, while at workshop three we introduced an experimental Retrieval Augmented Generation system that allowed people to ask questions directly of the data. This was used to critically explore potential applications of AI to identify risks, development needs, and opportunities.

For this paper, I intend to focus on a specific participant response: "Just chill". This phrase was said by a participant at the first workshop, and summarised a key thread of discussion. There was a clear sense of excitement for AI to undertake defined tasks with mutual benefits to libraries and the communities they serve – to assist with auditing collections, repatriation and restitution, automation of searches, identification of places and faces, reconnecting dispersed materials, and transcribing oral histories. Participants referred to "experimentation" with AI tools to understand their outputs and utility, and within private, encapsulated and controlled systems. Participants argued that not only should libraries give intellectual leadership of this process to Indigenous communities to ensure

concepts of self-determination and Indigenous data sovereignty (Maiam nayri Wingara and the Austraian Indigenous Governance Institute, 2018) are respected. This finding mirrored existing communiques such as the UN Declaration of the Rights of Indigenous Peoples (2007). However, in the context of AI systems design, a tension emerged between the relational modes of communication that would build trust, shared understanding, and consensus, and the fast-paced project lifecycles which underpin many implementations.

This tension between "fast" and "slow" ways of relating to technology has informed my own thinking on the implementation of AI in libraries. As the example of iREAL shows, source communities and librarians perceive the benefit of AI, and the related benefits of spending time to understand how it intersects with the life worlds of our user and source communities. This can produce understanding of how the technology operates and allows individuals to assert their right to reply. However, the idea of "fast" technology still holds relevance in an age where libraries fear being left behind the technological cutting edge. We can understand this moment through the literature that has emerged around "Slow AI" (Huggett, 2024), a term that echoes the slow technology movement originated by Hallnäs and Redström (2001). Slow technology is a design agenda for creating technologies that afford moments of mental reflection, not efficiency in performance, in recognition that "as the use of a certain kind of object changes there is often a need to reconsider the principles behind its design" (Hallnäs and Redström, 2001). As Hallnäs and Redström argue, the main purpose of information technology is to make people more efficient at carry out certain tasks, and this is a highly reasonable agenda. Fast technology, therefore, prioritises efficiency in functionality, with respect to a well-defined task. It minimises time spent with a technology by making the user more efficient and making the artefact easy and fast to use.

"Slow AI", though is about designing systems that encourage us to take time, think, reflect, and define the direction of travel (Huggett, 2024). It contrasts with fast technologies not because designing to save time is bad, but because as a critically engaged profession it is essential that we spend deep time with affected communities understanding the technologies we seek to adopt. Time, however, often feels like the thing we don't have, because of the concern of being left out of the aforementioned "big digital information party" (Liew, 2012). But what if we arrive at that party and we don't fit in? What if the people in attendance don't share our values? Do we just stay put and get stuck into the dancing, do we go home and pretend it ever happened, or is there a middle path where we manage a balance of "fast" and "slow" responses to technology?

Slowness, in the design sense, is not inaction or withdrawal from technological adoption. It is the creation of "time productive" (Hallnäs and Redström, 2001) engagements that involve us, our communities, and new technologies in moments of reflection that ensure we make AI work for us, and for our users. While I would caution against transposing Indigenous concepts of relationality into a Western context, the question of community self-determination is broadly applicable to the sector's approach to AI. If the library sector can take a nuanced approach to technical adoption, redefining fast adoption in ways that align with our values and those of our communities, and embracing slowness as a method that sits alongside and is productive of action, then this might help us to engage with the promises of AI in a manner that befits our profession.

For fast AI implementation, we might consider:

- Is it efficient, in that it saves time for us or our readers?
- Is it beneficial for us, our readers, and our source communities?
- Is its usage clearly defined and necessary?

For slow AI:

- Have we created space for reflection and enhanced understanding?
- Is what we're doing driven by, and responsive to, the needs and views of our stakeholder communities?
- Have we ensured that the time we spend with the technology is productive of usable outcomes?

For both, we need to ask fundamental questions about the ethics of our actions. Are we taking responsibility for the ways in which AI will be used in our institutions, and with our collections? If our actions enact harm, can that harm be easily undone? And are we including people in the room when we make decisions that might affect them?

Returning to McLuhan's idea of the "figure" – the technology itself – we don't need to panic about the technologies that come under the AI umbrella. We do need to take steps to work through the challenges, to identify its upsides, and to work with our user and source communities to co-create a future where we put it to work for the benefit of our staff, institutions, and communities. That requires work before we reach the planning and implementation phase, and it requires us to learn from past mistakes where the rich promise of a new technology was left unfulfilled because of actions, or inaction, that led to unintended barriers to our user communities. It also requires us to understand that technology itself is relational, and that we must embrace and uncover that relationality in our own work.

The Ground: AI Capitalism and Generative AI

Before wrapping up, I would like to dwell briefly on the "ground" on which this technology sits, because one of our greatest challenges for AI is going to be the way in which data capitalism has co-opted our sectors values and terminology to push forward business models that bear little resemblance to those values. Vincent Mosco (2004) argues that digital technologies fulfil the role of the sublime in contemporary society. Dating back to the writings of Immanuel Kant, sublimity refers to something awe-inspiring and overwhelming to our senses. This sense of overwhelm can give seemingly powerful new technologies a mythical status that obscures other aspects. Mosco, similarly to McLuhan, argues that we can only understand our own technological moment by looking at it both materially and culturally. AI, for instance, is both simultaneously subject to the rhetoric of innovation, and in need of careful unpacking to understand what is occurring underneath that rhetoric.

A major challenge for libraries in relation to AI is that the public discourse is dominated by major players who are perceived to be at the cutting edge of technological innovation. As Karen Coyle (2017) has noted on the development of the library catalogue: "if we look at the timeline of

information technology over the 20th century and into the 21st, we see library technology falling behind the general technology evolution." This point is more widely applicable to the portfolio of digital technologies, for which libraries are frequently adopters rather than creators. As Generative AI dominates the public discourse, the sums of money invested in Big AI companies show why their voices are so dominant. In March 2025, OpenAI – the creators of ChatGPT – undertook a $40 billion funding round to support the development of so-called 'Artificial General Intelligence' (Singh and Varghese, 2025). By comparison, the national libraries of Europe can each express their annual budget in the tens to low hundreds of millions. The direction of flow of capital tells us a lot, as Meredith Whittaker (2021) argues: "Modern AI is fundamentally dependent on corporate resources and business practices, and our increasing reliance on such AI cedes inordinate power over our lives and institutions to a handful of tech firms." However, transformation is hard to predict, and hype can often have a damagingly self-fulfilling aspect to them: "myths matter in part because they sometimes strive for their realisation whatever the cost" (Mosco, 2004). The hype around GenAI has built a sense of urgency for its adoption, and has had many of us so entranced by the figure that we risk ignoring the ground on which it sits.

EM Rogers' (2003) work on how innovations are diffused and adopted, differentiates between thought leaders – autonomous individuals that influence our understanding of technologies – and "change agents" – prominent individuals who look to influence the diffusion of new technologies in a manner that benefits their own agency. In looking at these change agents for Generative AI, we can see some emerging trends in their language that should make the library (Johnson and Refsum, 2025) community wary about the future. Such announcements focus on the idea that LLMs have ingested the "cumulative sum of human knowledge" (Milmo and editor, 2025), and intense lobbying efforts to adapt Intellectual Property laws to legalise extractive practices of scraping copyrighted materials from the open web (Dredge, 2025). Both have major implications for how the wider public, and the AI sector, will view library collections in the future. Moreover, we can see a strategic adoption of the language and role of the national library itself.

In January 2025, for instance, the UK government launched an AI Opportunities Action plan (Department for Science, Innovation and Technology, 2025a), with a press release on the same day utilising the language of intravenous drugs to claim "today's plan mainlines AI into the veins of this enterprising nation" (Department for Science, Innovation and Technology, 2025b). One of the suggestions in that action plan was the creation of a National Data Library, and the reaction of lobbyists to this prospect shows how the institution of the national library, and the role of the national librarian, risks being diluted. While libraries are envisaged in the Action Plan as a data source for creating a copyright-cleared media asset training dataset (Department for Science, Innovation and Technology, 2025a), there is little sense that the sector will be central to leading these efforts. Instead, the lobby response co-opts terms from the national library sector to effectively side line our contribution and expertise.

In early 2025, the Tony Blair Institute for Global Change released a report entitled "Governing in the Age of AI: Building Britain's National Data Library." One of their key recommendations for forming the National Data Library is that "National Data Librarians" should be embedded within key government departments and services. These "librarians" would be responsible for bridging the gap

between government data controllers and external users, and ensure each department has someone responsible for making data accessible. The vision for this role simultaneously evokes the tradition of librarianship, and rejects it out of hand:

> *While the National Data Librarian role draws inspiration from traditional librarian principles, it represents a distinct governmental function not governed by existing library accreditation frameworks. The role will not be defined by conventional library credentials and should be exempt from any associated regulation (Johnson and Refsum, 2025).*

This is where the greatest risk to the library sector with AI lies; not from the technology itself, but from powerful lobbyists who adopt our language and trade off our reputation as trusted custodians of information, but make it clear that they do not wish to be constrained by the checks and balances that define our profession. Our profession, and its implied values, are being used to sell a future imaginary for AI. We must therefore ask ourselves a fundamental question around data governance: who is in control of decisions about how library data will be used within, and for, AI? As Shannon Vallor (2023) argues in the Edinburgh Declaration on Responsible AI, utilising AI in a responsible manner requires us to accept responsibility for the implementation and consequences of AI in our institutions.

Conclusion

In my work, I try to understand what community involvement might do for developing better technologies in libraries, and how that can come into tension with the work of major national and research libraries which must address their regulatory responsibilities. However, those responsibilities are also in tension with data capitalism of the type that characterises Big AI. This form of AI adoption is global and hegemonic, not local and situated, and therefore will not account for the complex needs of our sector, or our users. Keeping track of both the technology, and the wider context, is a key responsibility for National Libraries as they navigate the global and local contexts for AI adoption. Finding a path through this challenging landscape does not mean ignoring AI: rather it means adopting technology in ways that support our sectoral mission, our organisational objectives, and our user and source communities. This will require each of us to attend to the full complexity of technological adoption. Watch the people, not just the technology. Attend to the ground, not just the figure.

References

| Bashir, F. and Warraich, N.F. (2023) 'Future libraries' blockchain opportunities and challenges: a systematic literature review and research agenda', Digital Library Perspectives, 39(3), pp. 293–310. Available at: https://doi.org/10.1108/DLP-08-2022-0065

| Bevan, N. (1994) 'Transient technology? The future of CD-ROMs in libraries', Program, 28(1), pp. 1–14. Available at: https://doi.org/10.1108/eb047155

| Bouich, A. et al. (2025) iREAL: Inclusive Requirements Elicitation for AI in Libraries to Support Respectful Management of Indigenous Knowledges. Literature Review. Available at: https://zenodo.org/records/15791074 (Accessed: 15 September 2025).

| Castells, M. (2002) The Internet Galaxy: Reflections on the Internet, Business, and Society. Oxford: Oxford University Press.

| Coyle, K. (2017) 'Creating the Catalog, Before and After FRBR'. Encuentro di Catalogacion y Metadatos, Universidad Nacional Autonoma de Mexico, 12 September. Available at: http://kcoyle.net/mexico.html (Accessed: 17 October 2017).

| Department for Science, Innovation and Technology (2025a) AI Opportunities Action Plan, GOV.UK. Available at: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan (Accessed: 15 September 2025).

| Department for Science, Innovation and Technology (2025b) Prime Minister sets out blueprint to turbocharge AI, GOV.UK. Available at: https://www.gov.uk/government/news/prime-minister-sets-out-blueprint-to-turbocharge-ai (Accessed: 15 September 2025).

| Dredge, S. (2025) 'Former Meta exec Nick Clegg makes "implausible" AI comments', Music Ally, 27 May. Available at: https://musically.com/2025/05/27/former-meta-exec-nick-clegg-makes-implausible-ai-comments/ (Accessed: 15 September 2025).

| Dudgeon, P. (2023) 'Caring for Country: Indigenous Wellbeing, Law, and Environmental Justice', in Climate Change and Mental Health Equality.

| Gooding, P. (2017) Historic Newspapers in the Digital Age: 'Search All About It'. Abingdon: Routledge.

| Gooding, P., Terras, M. and Berube, L. (2019) Towards User-Centric Evaluation of Non-Print Legal Deposit: A Digital Library Futures White Paper. Glasgow, Edinburgh and Norwich: University of Glasgow. Available at: elegaldeposit.org/resources (Accessed: 30 May 2019).

| Hallnäs, L. and Redström, J. (2001) 'Slow Technology - Designing for Reflection', Personal and Ubiquitous Computing, 5(3), pp. 201–212.

| Huggett, J. (2024) 'Slow AI', Introspective Digital Archaeology, 27 November. Available at: https://introspectivedigitalarchaeology.com/2024/11/27/slow-ai/ (Accessed: 28 January 2025).

| Indigenous Archives Collective (2021) Indigenous Archives Collective Position Statement on the Right of Reply to Indigenous Knowledges and Information held in archives. Available at: https://indigenousarchives.net/indigenous-archives-collective-position-statement-on-the-right-of-reply-to-indigenous-knowledges-and-information-held-in-archives/ (Accessed: 30 August 2023).

| Johnson, R. and Refsum, C. (2025) Governing in the Age of AI: Building Britain's National Data Library. Available at: https://institute.global/insights/tech-and-digitalisation/governing-in-the-age-of-ai-building-britains-national-data-library (Accessed: 15 September 2025).

| 'kingsdigitallab/iREAL' (2025). King's Digital Lab. Available at: https://github.com/kingsdigitallab/iREAL (Accessed: 15 September 2025).

| Lambert, S. and Ropiequet, S. (eds) (1986) CD Rom The New Papyrus: The Current and Future State of the Art. Microsoft Press.

| Liew, C.L. (2012) 'Towards Socially Inclusive Digital Libraries', in Digital Libraries and Information Access: Research Perspectives. London: Facet.

| Maiam nayri Wingara and the Austraian Indigenous Governance Institute (2018) 'Indigenous Data Sovereignty Communique', in. Indigenous Data Sovereignty Summit, Canberra, ACT. Available at: https://static1.squarespace.com/static/5b3043afb40b9d20411f3512/t/63ed934fe861fa061ebb9202/1676514134724/Communique-Indigenous-Data-Sovereignty-Summit.pdf (Accessed: 15 September 2025).

| McLuhan, M. (1967) The medium is the massage. Harmondsworth: Penguin.

| Milmo, D. and editor, D.M.G. technology (2025) 'Elon Musk says all human data for AI training "exhausted"', The Guardian, 9 January. Available at: https://www.theguardian.com/technology/2025/jan/09/elon-musk-data-ai-training-artificial-intelligence (Accessed: 15 September 2025).

| Mosco, V. (2004) The digital sublime: myth, power, and cyberspace. Cambridge Mass.: MIT Press.

| Office for Artificial Intelligence (2021) National AI Strategy. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf (Accessed: 15 September 2025).

| Rogers, E.M. (2003) Diffusion of Innovations. Fifth Edition. New York: Simon & Schuster.

| Singh, J. and Varghese, H.M. (2025) 'OpenAI to raise $40 billion in SoftBank-led round to boost AI efforts', Reuters, 1 April. Available at: https://www.reuters.com/technology/artificial-intelligence/openai-raise-40-billion-softbank-led-new-funding-2025-03-31/ (Accessed: 15 September 2025).

| Smith, L.C. (1976) 'Artificial intelligence in information retrieval systems', Information Processing & Management, 12(3), pp. 189–222. Available at: https://doi.org/10.1016/0306-4573(76)90005-4.

| The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (2013). Available at: http://www.legislation.gov.uk/uksi/2013/777/contents/made (Accessed: 15 September 2025).

| United Nations (2007) United Nations Declaration on the Rights of Indigenous Peoples. Available at: https://social.desa.un.org/issues/indigenous-peoples/united-nations-declaration-on-the-rights-of-indigenous-peoples (Accessed: 15 September 2025).

| Vallor, S. (2023) 'Edinburgh Declaration on Responsibility for Responsible AI', Medium, 14 July. Available at: https://medium.com/@svallor_10030/edinburgh-declaration-on-responsibility-for-responsible-ai-1a98ed2e328b (Accessed: 15 September 2025).

| Whittaker, M. (2021) 'The steep cost of capture', Interactions, 28(6), pp. 50–55. Available at: https://doi.org/10.1145/3488666.

| Wilson, T.D. (1981) 'On User Studies and Information Needs', Journal of Documentation, 37(1), pp. 3–15.

Richard Ovenden:
Burning the Books, Then and Now: From Clay Tablets to the World of AI

Attacks on knowledge have a long history. Examining that history – which I recount at length in my book "Burning the Books: A History of Knowledge Under Attack" (2020) tells us much about the importance of the preservation of knowledge and the place of institutions - libraries and archives - that society has entrusted that role to. At the end of this paper I want to bring the historical account up to date.

The library of Ashurbanipal, King of Assyria, dating from the 7th century BCE, and formed of clay tablets marked with cuneiform script is the best known of many libraries and archives that survive from the ancient civilisations of Mesopotamia, vital institutions dating back five millennia, that were also formed through acts of destruction and deliberate theft. The library survives, for the most part in the British Museum in London. There are accession records for Ashurbanipal's Library which have been studied by scholars show that he was deliberately targeting libraries and archives in neighbouring states - especially Babylonia – by sending his agents to go either forcibly or through diplomacy, to seize documents from these other libraries in order to build his own knowledge base up. Part of the content of these ancient libraries concerned the prediction of the future, about astronomy, astrology, and divination, a theme that recurs in the digital age: if you are able to remove knowledge from your enemy, you can not only make them weaker but you can make yourself stronger.

For millennia, the greatest library in the ancient world – in Alexandria - has been assumed to have been destroyed in a catastrophic conflagration. The ancient writers agree that it was larger than any other library they knew of, and that great scholars came to work there – such as Euclid the founder of modern mathematics – who wrote the Elements of Geometry, while working at the great library. What modern scholars agree on is that the library did not go up in flames in a single terrible event, but declined slowly, over a long period of time, reduced to nothing through neglect and under-funding, so that by the 4th century of the Christian era the library was completely gone, just a memory. The message for our own age is that we will lose libraries if they are not seen as a priority for modern nations.

One of the most momentous periods for the destruction of knowledge was the Reformation of the 16th century. One person – John Leland, was instrumental in both destruction and preservation of knowledge at the time. Leland was an astonishing character. Henry VIII tasked him with a 'most gracious commission, to peruse and most diligently search all the libraries' of the monasteries, searching for information to help him win his case for the divorce of Catherine of Aragon and to enable him to marry Anne Boleyn, and later to argue for the divorce of the whole country from Papal authority. In 1533 Leland visited Glastonbury Abbey. In size it was actually bigger even than Canterbury Cathedral, it attracted great wealth, many donations from pious pilgrims, but it also built up an extraordinary library. We don't know exactly how many there were in 1533-34 when Leland

visited, there were at least 1500-2000. A mere sixty volumes are known to survive today. A similar fate befell the other medieval religious libraries of Britain, none of whom escaped the Reformation intact. What happened to the books? From contemporary accounts we know that many of them were torn up and sold. Some were sold as scrap, only a few surviving through the actions of collectors known as antiquaries, whose personal collections became the basis of new institutional libraries, such as the Bodleian.

A similar fate befell the books in the University Library in Oxford. Originally founded in 1320, it grew during the middle ages through numerous gifts, especially a spectacular one in the middle of the 15th century. But this library was destroyed in the second phase of the Protestant Reformation, 1549-1550. Again, the books were mostly sold for scrap materials, and only a handful escaped with Catholics fleeing to Continental Europe. What followed was a reaction against this wholesale and ideologically driven destruction of knowledge. Sir Thomas Bodley, from a staunchly Protestant family, an Oxford graduate, and someone who had considerable private wealth, and he re-founded the library from 1598-1602 with a dual emphasis on both preservation of knowledge, and on access to it.

I'd like us to move forward now into the nineteenth century to another episode of the destruction of knowledge, the burning of the Library of Congress in 1814. This was the result of a military intervention led by Rear Admiral Sir George Cockburn, who led a British expeditionary force to the United States, a part of the war of 1812-14. There is an eye-witness account of the burning of Washington written by George Gleig who wrote: "I do not recollect to see more striking or sublime than the burning of Washington". He also was rather ashamed that the troops of which he was one also set fire to "a noble library, several printing offices and all the national archives which were committed to the flames, which might better have been spared", so he later admitted. The library was housed in the Capitol building, the only stone building in the city at the time, and which housed the Senate and the House of Representatives, as well as the office of the President, and the Library of Congress, the only library in the city. The Library of Congress in 1814 held 5,000 volumes, that provided a very useful set of combustible materials to start the fire.

What happened after the events of August 1814 was another response to destruction, and a further indication of that human impulse for preservation and renewal. That response came from Thomas Jefferson, one of the founding fathers of the United States, and a former President, who had retired to his estate at Monticello in Virginia, where he heard about the fire, and wrote a fierce letter to a newspaper in Washington saying that this was an act of barbarism and he offered his own library, the greatest private book collection in America at the time, to be purchased by Congress at favourable rates to replace the lost library. Congress eventually agreed to the purchase, allowing the Library to be renewed, starting it on the road to becoming the great institution that it is today.

Almost exactly a century later another noteworthy attack on knowledge which became an international incident was the destruction of the Library of the Catholic University of Louvain. In August 1914 the German troops set fire to the historic university library, which was quickly destroyed, with almost all the collections incinerated. The University library dated back as an institution to the 1630s, and was re-founded in 1835, becoming one of a number of legal deposit libraries for the (then) new country of Belgium. The attack on knowledge that the burning of the

library represented prompted an international outrage. All over the world the news of the burning of the library was met with outrage and horror. The burning of the library was viewed as a crime against the world, and the destruction of the Library of Alexandria was evoked to give a sense of the scale of the loss. But one of the interesting things about this story was the reaction to the great conflagration. An international movement to raise funds and to donate books to give to the library was begun. A special clause in the Treaty of Versailles was written whereby Germany was charged with replacing the destroyed books. The Americans took the library's renewal as an opportunity for projecting soft power in Europe after the First World War. A committee charged with raising the funds necessary to rebuild the physical structure of the library was formed, chaired by Nicholas Murray Butler, the President of Columbia University. Butler's Committee chose an American architectural practice, Warren and Wetmore, to design the new library, and they designed a pastiche or facsimile of the original building in the low countries vernacular style. A motto soon became associated with the promised building: 'Destroyed by the Germans in 1914. Restored by America in 1922.'

Despite the ambition it took Butler's committee much longer to raise the money than they had originally planned, with John D Rockefeller eventually supplying the shortfall himself. By the time that they finished raising the money, in the late 1920s, the post-war diplomacy between Belgium and Germany had begun to see a burying of the hatchet, so-to-speak, and the acts in Louvain in 1914 began to be purposefully ignored or downplayed by Belgians. The Americans intended a grand opening ceremony for the rebuilt library with a massive plaque laying out the motto in Latin - that the building was destroyed by the Germans, rebuilt by the Americans. This became a national point of tension. The American architect put this plaque up several times and local Belgians climbed up in the middle of the night and smashed the plaque because they didn't want it to colour the relations that they had with their neighbours. Eventually the plaque was removed and placed in a war memorial, and the library was finished: rebuilt and modernised. In 1940 the library was destroyed a second time, and again by the German army, who targeted artillery fire on the library. The Library was rebuilt after World War Two, for the second time.

The Holocaust was one of the episodes in history where the most destruction of knowledge takes place. Vilna, or modern day Vilnius, in Lithuania at the beginning of the twentieth century was one of the great centres of Jewish civilization in eastern Europe, a city full of books and archives, and learned Rabbis. The Strashun Library had been left by a bibliophilic Jewish businessman at the end of the 19th century to the Jewish community. On the eve of World War Two it had a busy reading room, and a learned librarian. But Vilna also had a great archival institution, a research institute into Yiddish culture, into the cultural life of everyday Judaism in Central and Eastern Europe, called YIVO. YIVO from its foundation in 1922 began to collect materials that documented everyday life of the Jewish communities in central and Eastern Europe.

In 1942 the Germans invaded and occupied Vilna, seizing the Jewish library and archive collections in the city. Just behind the Blitzkrieg came an operational group, established by and run by a librarian, Johannes Pohl. The group, called the Einsatzstab Reichsleiter Rosenberg, was tasked with identifying books and documents from the seized Jewish libraries and archives, which were to be sent back to Frankfurt, to the 'Institute for the study of the Jewish question' established by Alfred Rosenberg, the chief architect of Anti-Semitism in Nazi Germany. What was not chosen to be sent to Frankfurt would

be sent to local paper mills for destruction. The Nazis forced the Jewish community of Vilna to live in the ghetto, and they identified a number of former librarians and archivists and other intellectuals to undertake the horrible task, at gunpoint, of sorting through these seized libraries and archives. Such a task would have been incredibly difficult with their own history and culture either being sent to Germany, or to be destroyed. The Jews who were selected for this task became known as the 'Paper Brigade'.

The human impulse toward preservation can be identified in the actions of the 'Paper Brigade'. They smuggled items from the collections they were forced to sort through back into the ghetto every day, and they hid these books and documents inside the ghetto itself, in the hope that one day they could be recovered. Each time they did this, they risked their own lives, displaying a compulsion to preserve their own culture, their own documentary witness to their community, to their civilisation, in the hope that they would survive the Holocaust and the documents could speak of the lives they had before the war. A few of the members of the Paper Brigade managed to escape when the Vilna ghetto was liquidated in 1944 and joined Partisans in the forests. With the Soviet army they liberated Vilna and retrieved some of the collections - tens of thousands of documents that they had managed to hide.

This effort to preserve the documentary heritage - witnesses of Jewish life – did not just happen in Vilna, it was undertaken in other centres in Eastern Europe as well. In the Warsaw ghetto an archive was made by an organisation called 'Oyneg Shabes', led by an extraordinary man called Emmanuel Ringelblum, who was murdered when the Warsaw Ghetto was liquidated, but only after he had managed to hide and bury documents which he and his fellow members had saved. These were dug up afterwards in metal cartons and milk canisters.

Some of the documents which had found their way from Vilna to Rosenberg's Institute in Frankfurt were seized by American forces in 1945 and were eventually sent back to a branch of the YIVO Institute in New York in 1947. Meanwhile, back in Vilna, the materials that had been saved by the 'Paper Brigade' and then sent for destruction again by the Soviets were saved a further time, this time by a Lithuanian librarian called Antanas Ulpis. Ulpis preserved these documents by going to the paper mills and turned the trucks around. He hid them in a church that had been requisitioned as one of the storage sites for the new National Library of Lithuania. Ulpis hid the documents away in organ pipes in other locations, and they only became revealed after Ulpis's death in 1989, as the iron curtain came down. They are now one of the great treasures of the National Library of Lithuania and are being digitised by the YIVO Institute in New York.

The attacks on knowledge in Bosnia and Kosovo during the Balkan conflicts following the break-up of the former Yugoslavia is another example of a cultural genocide that came before a human genocide. The National Library of Bosnia and Herzegovina in Sarajevo was deliberately attacked by Serb militia besieging the city with incendiary shells. No other buildings were targeted on that day, August the 25th, 1992. The fire fighters and librarians who tried to rescue collections from the burning building were shot at by snipers. The library was attacked because it was a living institutional symbol of the multicultural community that Sarajevo and Bosnia had managed to create in the decades before the wars. The library preserved the written culture of Bosnian Muslims, Jews, and Christians

all living together, but something which the Serbs deliberately sought to attack through eradicating the Bosnian national memory – the library.

Over the last five years, since the publication of "Burning the books", a whole series of attacks on knowledge have taken place which show the desire for authoritarian regimes to exercise control over knowledge. In August 2021 the Taliban rapidly took control in Afghanistan, destroying a number of libraries, especially the few that had been established to support the education of women and young girls. From February 2022 we have seen the widespread destruction of libraries and archive in Ukraine, especially university and public libraries, and in Russian-occupied Ukraine books in Ukrainian are being destroyed by Russian forces. In America from 2022 there has been an epidemic of book banning in many US states, targeted mostly at local school and public libraries, led mostly by right wing groups intent on limiting the freedom of library patrons to shoes what to read, and using libraries and librarians as targets in the culture wars.

In February 2025 Trump's attempt to control the archive took a decisive step with the dismissal, on spurious grounds of the National Archivist. This was followed by the removal of the members of the Committee advising on the classification of records. In April the Librarian of Congress was also dismissed, together with the head of the Copyright Office. Again on spurious grounds, alleging that the Librarian was providing inappropriate books to children (even though you have to be 16 to be admitted to the library). The Trump administration also unleashed a programmatic deletion of content from federal government websites, including over 180k pages from the website of the Centre for Disease Control. Here again the age-old impulse to preserve knowledge has taken the more modern form of initiatives such as the Data Rescue Project, who have been preserving websites as a kind of 'digital rescue' mission.

The rise of AI opens up further challenges for libraries, and shows further signs of domination in the digital regimes of knowledge. The first of these has come as the companies reach 'the data frontier', as the training sets for Large Language Models have exhausted the open web, they are looking for more words. In copyright books have been part of their attempt to deal with this problem, with many companies engaging in piracy (and one of them, Anthropic, settling out of court). At the Bodleian we are engaging with a pilot project too with with OpenAI. My view is that libraries cannot ignore AI, and must engage seriously with the industry. We are testing the use of the technology in workflow design, in data exchange, and in testing automated processes against human processes. We will share the output of our project when it is complete (in 2026).

One other aspect of AI is challenging a long-held aspect of librarianship – openness. The company that builds the training sets for AI companies by harvesting the open web (Common Crawl) publishes the list of its most frequently crawled domains. You can see from this list that it ranks university websites very highly, and the Oxford university website highest among them. Much of this crawling is to extract data from theses. Is this a good thing for universities like Oxford? One view is that this surfaces Oxford content in the outputs of generative AI – making the trusted knowledge from Oxford appear in the outputs of these tools. On the other hand, the AI companies are ruthlessly capitalist and are exploiting this free content for commercial gain. Do we not owe a duty of care to recent doctoral students, whose theses data is being exploited in this way? In other words, should

universities continue to be open with all of their knowledge online? These are going to be major issues into the future.

At the end of the day, libraries still have some of their brand values intact. They are trusted sources of information, much of this derived from the fact the libraries have a track record of preserving knowledge and safeguarding it for future generations. In the digital age we must hold onto these values and ensure that they are not eroded.

Martin Öövel:
Ensuring Linguistic and Cultural Diversity: AI – Ally or Enemy?

Introduction

Libraries are guardians of knowledge, but they are not – and never have been – the only ways to access it. One way to conceptualise the historical development of information access is in four stages:

1) the *messenger*, slow but transparent, where the source is known and trusted, much like a modern expert speaking on television today;
2) the *library*, offering curated diversity of sources, reliable though slower to navigate;
3) *internet search*, faster but with quality varying and results curated by opaque algorithms;
4) and finally *large language models (LLMs)*, the fastest yet least transparent, producing single-output answers with hidden origins.

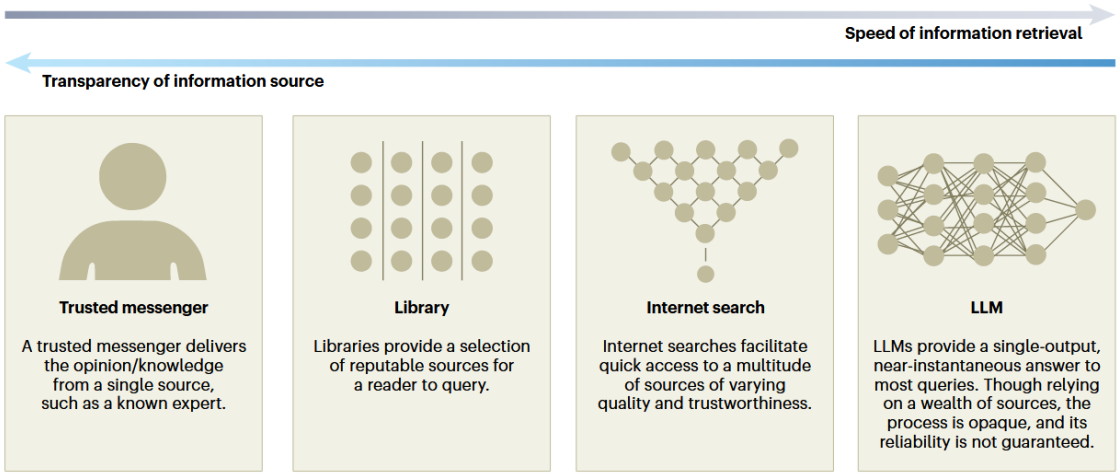


Figure 1. Ways of information Access. Burton, J. W. et al. (2024). How large language models can reshape collective intelligence. Nature Human Behaviour, 8(9), 1643–1655. https://doi.org/10.1038/s41562-024-01959-9, Last visited: 5 December 2025

This trajectory can be framed as a trade-off between speed and transparency. Now that access to knowledge has become more convenient than ever thanks to AI, the crucial question is: transparent for whom, and at what cost? Transparency is not just about trust, but also about language, culture, and who gets to participate in the knowledge landscape.

Changing Conditions for Libraries

The primary responsibility of libraries has always been to connect people with knowledge. However, the rapid emergence of powerful AI technologies is reshaping this function. AI is already in use as personal assistants, in classrooms, and in national deployments such as those in Estonia, Saudi Arabia, and Sweden. Individuals now have personal AI-powered librarians and educators accessible

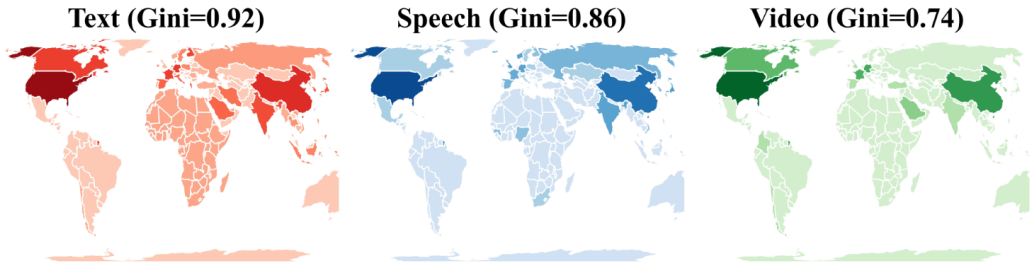
via their devices. The shift is evident: in 2025, ChatGPT began to outpace Wikipedia in total monthly visits.

This turn toward speed and convenience, but away from transparency and traceability, demands serious attention from libraries. Libraries are centres of verified, inclusive, and contextual knowledge. When AI tools obscure their sources and reflect a narrow cultural viewpoint, the risk is not only inaccuracy but also the gradual erasure of smaller languages and cultures from the digital knowledge ecosystem.

I argue that libraries must take a leading role in ensuring that the development and use of AI protects linguistic and cultural diversity by shaping how knowledge is represented, accessed, and preserved across all languages.

Centralisation of AI

AI is technologically and culturally centralised. AI advancements are primarily driven by large technology companies operating predominantly in English and Mandarin, languages with substantial global market appeal. These corporations show limited commercial interest in supporting smaller languages, which represent only a fraction of the training data in widely used AI models. The result is striking: A Gini coefficient of 0.92 in dataset distribution, far exceeding inequality levels found in societies, showing that a handful of languages overwhelmingly shape AI development.



	Africa	Asia	Europe	N America	Oceania	S America
			By Count			
Text	0.3	13.4	24.0	61.5	0.7	0.2
Speech	3.6	35.7	30.4	30.4	0.0	0.0
Video	0.0	25.2	24.4	48.0	0.8	1.6
			By Tokens or Hours			
Text	0.0	6.1	55.4	38.4	0.1	0.0
Speech	0.1	38.8	18.8	42.4	0.0	0.0
Video	0.0	23.1	22.0	38.2	16.7	0.1

Figure 2. Gini coefficient of dataset distribution. Longpre, S. et al. (2024). Bridging the Data Provenance Gap Across Text, Speech and Video (arXiv:2412.17847). arXiv. https://doi.org/10.48550/arXiv.2412.17847, Last visited: 5 December 2025

Language, Culture, Fairness, and Knowledge Quality

The issue is not only about language, but also about culture, fairness, and knowledge quality. Performance gaps are well documented. A 2025 study at University College Cork found that leading LLMs scored over 20 percentage points lower in Irish than in English when answering the same exam-style questions. Similar patterns appear in other languages and cultures. Multilingual safety research further shows that models are less reliable and more vulnerable to harmful outputs in non-English languages. These systems are not culturally neutral, but instead reflect the biases of their data.

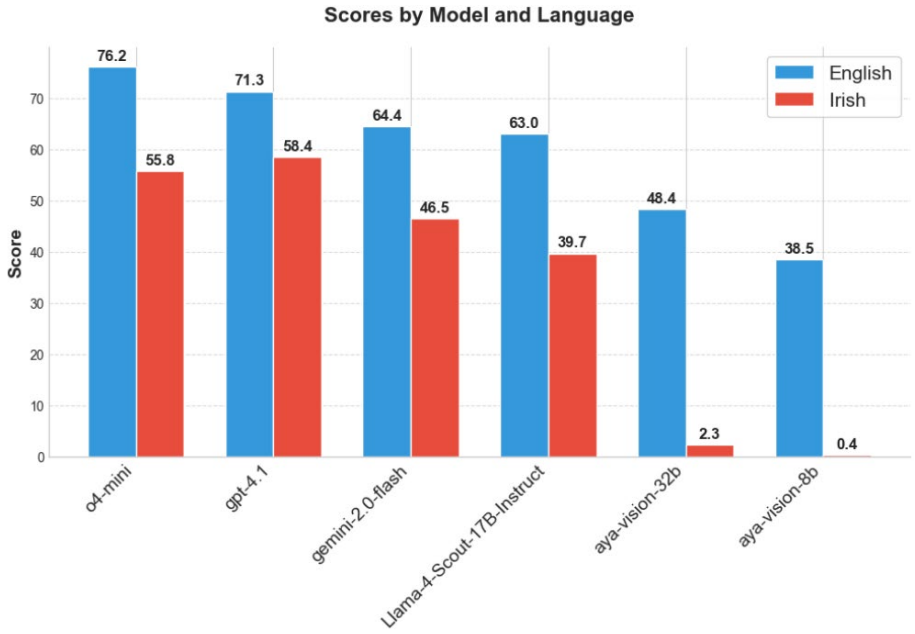


Figure 3. LLM performance in English vs Irish. Tran, K.-T., O'Sullivan, B., & Nguyen, H. D. (2025). IRLBench: A Multi-modal, Culturally Grounded, Parallel Irish-English Benchmark for Open-Ended LLM Reasoning Evaluation (arXiv:2505.13498). arXiv. https://doi.org/10.48550/arXiv.2505.13498, Last visited: 5 December 2025

Libraries must also advocate for inclusive copyright frameworks. For many smaller languages, much of the available data is copyrighted. Without ethical pathways for its responsible use, there is no realistic way to achieve parity.

The Role of Libraries

Libraries are uniquely positioned to address these issues. They can advocate for better representation in AI and multilingual standards, curate diverse, high-quality content for training datasets, and influence policy by demanding transparency and inclusion in AI development. They also have a vital role in shaping copyright frameworks. As the International Federation of Library Associations (IFLA) has emphasised, copyright should not be used as a blunt tool to solve unrelated ethical concerns. Importantly, these efforts are most effective when pursued collectively – through CENL, through collaboration, and through shared infrastructure.

Only culturally aware AI can truly be an ally for libraries. If AI lacks cultural and linguistic nuance, it cannot fully support librarians' work, whether regarding metadata, subject classification, or user engagement. At our own library, we are developing an AI-powered automatic cataloguing solution, a project we are proud of. Yet its performance is constrained by the underlying models, which are not trained with our specific linguistic and cultural context in mind. This limitation sets a ceiling on what the system can reliably achieve.

The Problem Will Not Solve Itself

One counterargument is that AI will inevitably become multilingual. This is not necessarily the case. Progress in smaller languages is often incidental, not intentional, and medium-resource languages risk being misled by partial success, believing that "ChatGPT speaks [language x]," when in fact its performance is shallow and unreliable.

Another claim is that market solutions and open-source communities will solve the problem. Yet markets prioritise scale over inclusion, and open-source communities, while valuable, still depend on data and infrastructure that many languages lack. Without institutional leadership, these gaps will persist.

Finally, some argue that the issue is too technical for libraries. On the contrary, libraries do not need to build AI themselves, but they must understand it, guide its use, and shape its impact. Dedicated expertise for AI, just as already exists for digital preservation or metadata, is essential for fulfilling the mission of libraries in our age.

Conclusion

Is AI an ally or enemy? It is neither. It is a tool for accessing knowledge, and thus inhabits the very space where libraries have traditionally existed. The more pressing question is in whose hands this tool works best: whose knowledge it amplifies, whose language it speaks, and whose stories it forgets.

Libraries are the right institutions to ask these questions, but also to help provide answers. If libraries lead, AI can become a true ally—not by replacing them, but by extending their reach, reinforcing their values, and preserving the richness of linguistic and cultural diversity for generations to come.

Wilma van Wezenbeek:
Responsible AI in the National Library's Strategy: Balancing the Social, the Digital, and the Green

KB, National Library of the Netherlands works towards a smarter, more skilled and more creative Netherlands.

KB has been a source of inspiration and development for centuries. Since its foundation in 1798, the focus has been on making the National Library collection visible, usable and sustainable for all Dutch people, for any purpose, both now and in the future.

Value to society is central: as a publicly financed organisation, the contribution we can make to society is paramount to us. To this end, we constantly monitor the progress and success of all our activities.

Digital at the forefront: in order to effectively fulfil our role in the information society, we are increasingly learning throughout the organisation to think and work primarily digitally.

Thinking in networks: our social role in the information society also includes thinking in networks. We are joining forces with users and partners to ensure continued progress.

Introduction: AI at a Turning Point

Artificial Intelligence has moved beyond laboratories and tech companies. It has become a system technology, across industries and governmental organisations, comparable to electricity or the combustion engine, that will reshape societies at their core.[1] National libraries, with their dual mission of safeguarding cultural heritage and enabling public access to knowledge, now stand at a critical juncture.

On the one hand AI offers extraordinary opportunities: collections that were once hidden in basements or bound by fragile paper can be brought to life, made discoverable, and connected in new ways. On the other hand, it raises fundamental ethical, legal, and ecological questions. How do we use AI in ways that protect the rights of authors, respect the privacy of end-users, and safeguard the environmental sustainability of our digital infrastructures?

The argument of this article is straightforward but urgent: the responsible use of AI requires balance, between social values, digital innovation, and environmental sustainability. Just as the open access and open science movements have taught us, this is not something individual institutions can achieve in isolation. It requires collective effort, built on trust, collaboration, and a shared vision.

[1] Mission AI. The New System Technology (2021). WRR / Netherlands Scientific Council for Government Policy. https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie, Last visited: 5 December 2025

The Social Dimension: Caring for Communities

For centuries, libraries have been custodians of public values. Neutrality, inclusiveness, accessibility, and reliability have been the cornerstones of their legitimacy. In the age of AI, these values are more relevant than ever.

AI systems affect the way people search for information, encounter knowledge, and form opinions. When used correctly, people can find information, develop themselves and participate in society. But left unchecked, they can amplify existing inequalities: biased algorithms can disadvantage vulnerable groups; opaque recommendation systems can deepen filter bubbles; generative tools can spread disinformation at unprecedented scale. National libraries cannot afford to stand by. They must function as counterweights, ensuring that technology remains aligned with the values of the communities they serve. Implementation of the EU's AI Act and Digital Services Act is one way to achieve this.

Practical examples already exist too. In the Netherlands, the AI Parade[2] brought AI awareness into public libraries across the country, reaching more than a million citizens. Through workshops, exhibitions, and playful interactions, people of all ages engaged with AI not as an abstract phenomenon, but as a tangible part of daily life. This initiative showed that libraries can be trusted guides in helping communities understand and question novel technologies.

One vivid example comes from the AI Parade. In a local library in Rotterdam, children were invited to interact with an AI tool that could generate poems based on their names.[3] For many, it was the first time they realised that machines could create something that felt personal and creative. Parents, on the other hand, asked critical questions: Who owns these poems? What happens to the data we feed into such tools? These moments of curiosity and critical reflection illustrate how libraries can encourage dialogue and bridge generations in exploring AI.

The Digital Citizenship programme[4] builds upon this role. By embedding AI awareness into broader efforts on digital inclusion, libraries ensure that citizens of all ages, not only the tech-savvy, gain the skills to navigate digital society. The AI Inspiration Guide[5] developed for public libraries offers concrete programming ideas: from workshops on detecting deep fakes to intergenerational dialogues where teenagers explain algorithms to older visitors. Such activities make AI tangible and support the development of critical digital citizenship.

---

[2] https://muzus.nl/en/projecten/everyday-intelligence-ai-parade/, Last visited: 5 December 2025

[3] https://poembooth.com/, Last visited: 5 December 2025

[4] https://www.bibliotheeknetwerk.nl/digitaal-burgerschap (in Dutch), Last visited: 5 December 2025

[5] https://www.bibliotheeknetwerk.nl/sites/default/files/documents/AI-inspiratiegids.pdf (in Dutch), Last visited: 5 December 2025

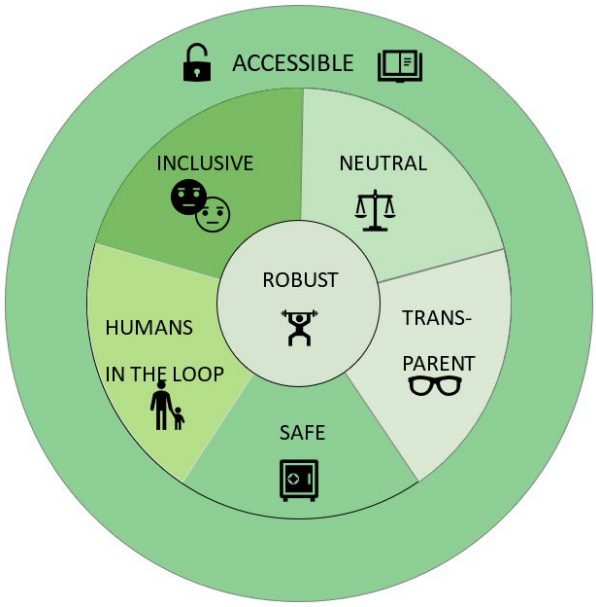


Figure 1: The principles for using AI in libraries help to ensure that technology is used responsibly within our domain. A new version is under development. Image from Wessel, J.W. (2020). https://doi.org/10.5281/zenodo.3865344

Yet raising awareness is only the first step. Libraries must also embed responsible AI practices in their own operations. That means upholding principles of transparency and explainability, ensuring that metadata generated by AI is clearly identified, and maintaining humans-in-the-loop for critical decisions. It also means collaborating with educators, policymakers, and civil society to promote AI literacy as a cornerstone of digital citizenship.

In short, caring for communities in the AI age means more than offering access to books or databases. It means helping people navigate a digital society where machines increasingly mediate what they see, know, and believe.

The Digital Dimension: Stewardship of Collections

National libraries are witnessing an unprecedented surge in digital content. Born-digital publications, digitised archives, and multimedia collections are growing at such a pace that infrastructures struggle to keep up. AI is often seen as the solution, capable of enriching metadata, supporting full-text search, and enabling conversational access through chatbots.

And indeed, AI brings remarkable potential. Experiments with language models show how visitors can interact with collection objects in natural language. Tools for automated metadata generation and retro-digitisation accelerate access to previously hidden sources. Responsible recommenders, developed in collaboration with computer scientists, demonstrate how AI can guide users without undermining diversity or neutrality.

Concrete projects highlight both the promise and the dilemmas of AI in collections. CuratorBot, developed with Delft University of Technology, allowed visitors to 'talk' to an object from KB's special

collections.[6] Asking questions like "What was life like in 17th century Amsterdam?" suddenly produced answers voiced through a chatbot in the persona of a historical artefact. While playful, it sparked serious reflection on authenticity and interpretation: was this the voice of history, or a modern construct trained on curated data?

Figure 2: CuratorBot, the result of a pilot by Delft University of Technology and KB, has given libraries an insight into how users can interact with and learn more about the collections in the near future.

Another project, developed with the Dutch national research institute CWI, focused on creating responsible recommenders for public libraries. Unlike commercial platforms that optimise for clicks, these recommenders were designed to balance diversity, relevance, and neutrality. They demonstrated that AI can support discovery without reinforcing echo chambers.[7]

Equally transformative – but with focus on internal processes – is the Retrotool project, which automates the generation of cataloguing data and the de-duplication of donated book collections.[8] By combining machine learning with human oversight, it accelerates access to millions of pages that would otherwise remain invisible. Yet each of these projects also raises questions of trust: how transparent must we be in labelling machine-generated metadata, and how do we ensure that human curators remain in control?

Policy responses are just as crucial. In 2024, KB released its AI statement on the commercial use of AI, defining clear boundaries for re-use of digitised collections. [9] In addition its collaboration with other partners in the GPT-NL project[10], a government-funded initiative to develop a transparent Dutch large language model, these efforts illustrate how libraries are actively contributing to the AI-driven digital landscape rather than merely responding to its development.

---

[6] https://visboeck.vercel.app/, Last visited: 5 December 2025

[7] See below: Acting collectively

[8] https://www.kb.nl/en/over-ons/projecten/automatisch-metadateren, Last visited: 5 December 2025

[9] https://www.kb.nl/en/ai-statement, Last visited: 5 December 2025

[10] https://gpt-nl.nl/ (in Dutch), Last visited: 5 December 2025

Stewardship in the digital dimension means more than managing servers and metadata. It means being ethical gatekeepers: enabling innovation, while setting boundaries that protect the integrity of knowledge and the trust of society.

The Green Dimension: Caring for the Planet

When discussing AI, conversations often focus on ethics and law. Yet one dimension is sometimes overlooked: environmental sustainability.

Digital infrastructures are energy-intensive: every digitised collection stored in a server park, every AI model trained on millions of documents, every query processed by a large language model consumes electricity and leaves a carbon footprint. For national libraries, institutions committed to serving both present and future generations, this is a sincere concern.

The ReVerDi project has pioneered methods for assessing library infrastructures through the lens of sustainability.[11] By applying life-cycle analysis, it examines not only the economic cost of storage solutions, but also their social and ecological impact. The results are sobering: even well-meaning digitisation projects can carry hidden environmental costs if not designed carefully.

The ReVerDi project makes these issues more visible. By comparing the full life cycle of physical versus digital preservation strategies, it revealed that digital solutions are not automatically 'greener' than traditional solutions. For example, long-term cloud storage of high-resolution scans can generate higher emissions than carefully managed physical archives. Conversely, digital access reduces travel-related emissions by allowing users to consult sources remotely. The key is to weigh trade-offs transparently.

One anecdote illustrates this tension. A historian once described how travelling across Europe to consult rare newspapers was both costly and carbon intensive. With digitised versions available through platforms like Delpher[12], research became faster and greener. However, this convenience is supported by an infrastructure of which energy consumption must also be considered. National libraries are uniquely positioned to weigh such trade-offs and to model sustainable digital infrastructures for the wider cultural heritage sector.

Taking the green dimension seriously means recognising that access and preservation are not limitless. Choices must be made, and sustainability must be part of those choices from the start. By embedding environmental criteria into their AI strategies, national libraries demonstrate that cultural heritage institutions can be leaders in balancing digital innovation with planetary responsibility.

[11] https://delftdesignlabs.org/projects/reverdi-real-versus-digital-reverdi-sustainability-optimization-for-cultural-heritage-preservation-in-national-libraries/, Last visited: 5 December 2025

[12] https://www.delpher.nl/, Last visited: 5 December 2025

Acting Collectively: The Power of Networks

However, one national library cannot resolve these challenges by itself. The scale of AI, its technological complexity, its legal uncertainties, its ecological footprint, requires collaboration. Fortunately, libraries have a long tradition of working together.

Networks such as CENL, IFLA, and AI4LAM already provide platforms for exchanging expertise, aligning strategies, and developing shared frameworks. Research collaborations, such as the Cultural AI Lab and HAICu[13], bring together computer scientists, humanities scholars, and heritage professionals to co-create responsible AI tools.

These collective efforts are more than practical necessities; they are expressions of a philosophical approach. Just as universities and their libraries collaborated to achieve open access and open science, resulting in transformative agreements, responsible AI in libraries must be built on shared norms and mutual trust.

Collective projects across Europe demonstrate what is possible. The EURIDICE project, funded by the EU, supports libraries in developing maturity models, training staff, and building capacity across the public library sector.[14] By pooling resources, national libraries ensure that even smaller institutions can benefit from shared expertise.

Figure 3: To ensure fairness and inclusion in the digital transition and to overcome the Digital Divide, Europe needs all-round digitally skilled people. Therefore, the EU funded project EURIDICE aims to educate and build 'social digital innovators', intensively trained to understand the nexus of society and technology, equipped with advanced digital skills and social reflexivity, and thus capable of driving human-centred innovation in inclusive and responsible ways in a rapidly changing society. A joint master's program will start at three universities in the fall of 2025. This image depicts a fair and inclusive digital society – a sustainable development goal that does not yet exist – but hopefully EURIDICE will contribute to.

The Cultural AI Lab provides another inspiring case. Here, computer scientists and heritage professionals jointly explore how AI can enrich humanities research while embedding ethical reflection from the start. One PhD project, co-financed by KB, is pioneering responsible recommender systems specifically for the library sector.[15] These initiatives show that collaboration not only shares risks and costs but also creates entirely new forms of knowledge.

---

[13] https://www.cultural-ai.nl/, Last visited: 5 December 2025; https://www.haicu.science/, Last visited: 5 December 2025

[14] https://euridice.eu/, Last visited: 5 December 2025

[15] https://www.cultural-ai.nl/projects/responsible-recommenders-in-the-public-library-sector, Last visited: 5 December 2025. Follow-up project DiBiLi (2026) should enlarge insights in potential biases in recommender

Looking beyond the Netherlands, the AI roadmap published by the Bibliothèque nationale de France in 2021 has inspired peers across Europe.[16] It demonstrated how ethical principles can be translated into concrete frameworks, milestones, and policies. KB is now preparing a similar roadmap, not just for itself but as a contribution to the broader CENL network.[17]

Finally, the debates encouraged by KB's decision to restrict AI crawlers from its digital platforms in 2024 underline why shared frameworks are essential. Heritage institutions found themselves caught between creators who demanded stronger protection and researchers who pleaded for open access. These tensions cannot be resolved by any single library. Only collective agreements, similar to those that enabled open access, can provide clarity, legitimacy, and balance in the AI era.

Acting collectively means recognising that our challenges are shared, that stepping over one's own shadow is required from time to time, and that our solutions will only be credible if they are built together.

Call to Action: Balancing the Future

The path ahead is clear, though not easy. AI is reshaping the very conditions under which libraries operate. To remain relevant and trustworthy, national libraries must embrace AI responsibly. That requires:

- **Social commitment**: upholding public values, protecting citizens, and fostering AI literacy
- **Digital stewardship**: managing collections ethically, setting boundaries for data use, and developing transparent infrastructures
- **Environmental responsibility**: embedding sustainability into every digital and AI initiative
- **Collective action**: aligning strategies, sharing knowledge, and building common roadmaps

The stakes are high. Without collective commitment, libraries risk becoming passive data providers in an AI ecosystem dominated by commercial or geopolitical interests. With commitment, however, they can shape AI in ways that serve communities, protect culture, and care for the planet.

Responsible AI is not an optional add-on. It is the essence of what national libraries stand for: ensuring that knowledge, culture, and values are preserved and passed on, ethically, digitally, and sustainably

systems for libraries. https://www.cwi.nl/en/news/dashboard-to-expose-bias-in-library-ai-systems/, Last visited: 5 December 2025

[16] https://www.bnf.fr/en/artificial-intelligence-bnf, Last visited: 5 December 2025

[17] In the meantime, the way KB deals with AI can be found here: https://www.kb.nl/en/over-ons/expertises/artificial-intelligence, Last visited: 5 December 2025

Gilles Pécout:
AI in Libraries: An Ethical Approach

Executive Summary of Gilles Pécout's intervention at the 39th CENL Annual General Meeting

The rise of AI represents a revolution for the cultural sector, especially for libraries. With strong digitalisation policies and the constitution of digital collections, libraries have built vast reservoirs of patrimonial data enriched with high-quality metadata.

In 2021, the Bibliothèque nationale de France (BnF) adopted an AI roadmap outlining its strategic and operational directions, defining AI as a tool serving its core missions: cataloguing, preservation, dissemination, accessibility, and discoverability of collections. Since, the emergence of generative AI has raised new questions, notably regarding data provision for training large language models (LLM) and the status of AI-generated creations.

These technological developments bring forth ethical challenges that libraries must address:

- *Support of professionals*: AI is likely to modify tasks associated with certain professions. Libraries must anticipate these changes and support their staff through awareness-raising initiatives and training on the opportunities AI offers.

- *Trust and rights:* As part of their legal deposit mission, libraries serve as trusted third parties for authors, publishers, and producers. They must uphold copyright and intellectual property rights, regulate partnerships with private AI actors, and help develop fair remuneration models for rights holders.

- *Data diversity and bias prevention:* Libraries must ensure data representativeness, implement bias correction mechanisms, and promote AI model transparency, explainability, and auditability. They should contribute to LLM training with diverse cultural data in national languages and support international cooperation on linguistic diversity, such as Francophonie initiatives.

- *Ethical AI research:* AI is a field of research like any other. As institutions supporting and contributing to research, libraries must uphold ethical principles in both their research activities and their partnerships. They must particularly promote open research.

- *Digital literacy*: Libraries must play a role in raising public awareness and educating citizens about AI tools, helping them navigate the new informational landscape shaped by these technologies.

- *Sustainability and international solidarity*: Libraries must minimise their ecological footprint by addressing the energy consumption of digital infrastructures and the massive storage of

data. Furthermore, they must contribute to shaping geopolitics of international solidarity focused on digitalisation and AI.

Addressing these challenges requires a coordinated European approach. The AI group of the CENL is drafting a common ethical charter, to be presented this year, aiming to harmonise national libraries' practices across Europe. This initiative reaffirms European libraries' commitment to aligning technological innovation with ethical principles in the use of cultural patrimony for AI.

Anna Chulyan:
National Library of Armenia - Partnership with Private Actors for Leveraging AI Technologies

The preservation of Armenian written heritage is both a cultural and social responsibility and a technological challenge. The digitisation of historical documents is considered to be of fundamental importance to modern archival practice. However, a significant number of non-Latin scripts, particularly those with underdeveloped digital infrastructures, continue to encounter substantial challenges. From medieval manuscripts in Classical Armenian (Grabar), to nineteenth-century newspapers and modern printed books, the Armenian language embodies layers of history that are invaluable for scholarship and national memory.

The National Library of Armenia (NLA) is the custodian of a vast collection of printed materials, including rare books and periodicals, totalling over eight million items in various languages. The National Library of Armenia is the largest repository of Armenian printed heritage in the world. The library began digitising its holdings for scholarly and public access more than a decade ago. The digitised content, comprising over 15 million pages, is presented comprehensively in the www.nla.am[1] databases.

Digitisation and Optical Character Recognition (OCR) for Armenian Heritage: Challenges and Perspectives

The Armenian alphabet, with its unique graphemes, ligatures, and orthographic evolutions, poses significant challenges for generic optical character recognition (OCR) tools. Although digitisation projects have already made significant progress in safeguarding fragile originals, true accessibility depends on OCR and handwritten text recognition (HTR) technologies. These technologies can transform static images into searchable, analysable text. However, applying OCR and HTR to Armenian heritage presents a complex set of obstacles spanning linguistics, technology, infrastructure, and institutional readiness.

The diversity of Armenian writing systems poses a significant challenge. There are three major traditions in the Armenian language: Classical Armenian (Grabar), Western Armenian, and Eastern Armenian. Each tradition has its own orthographic rules and historical spellings. It has been demonstrated that OCR systems that have been trained on one variant often encounter difficulties when confronted with another. The Armenian script, which is rich in diacritics and ligatures, requires a higher level of precision than Latin-based alphabets. Additionally, a significant proportion of heritage sources are characterised by multilingualism. In the context of newspapers and journals, for example, Armenian text was often mixed with Latin, French, Persian, and Russian. The presence of a mixture of scripts in this environment poses a significant challenge to automatic recognition.

[1] Last visited: 5 December 2025

This is because OCR engines must accurately identify and switch between languages on a single page.

The fonts used by Armenian printing presses reflect the distinctive features of the Armenian script. The first Armenian book was printed in Venice in 1512–1513 by Hakob Meghapart, the first Armenian printer. Later, other publications followed in various cities around the world. The reproduction of the fonts reflects the typographical peculiarities of different periods and countries. They may differ from one another in terms of size, form and arrangement. Furthermore, early printed books use glyphs that are absent from modern Unicode sets.

In the year 2021, by the Armenian Portmind company development of character recognition software for the Armenian language was undertaken. The Armenian document OCR solution also supports Latin and Cyrillic characters. It is designed to work with scanned documents. It can handle documents with different layouts, densities, and scan qualities. The software was specifically created for digitising printed content from Armenian books and other continuing resources (linear printed text). In collaboration with library specialists, an experiment was conducted over a period of eight months. In 2022, a memorandum of understanding (MOU) was signed between the National Library and Portmind. Currently, around 30% of the library's digitised databases use character recognition[2].

During the OCR process using Portmind's software, we encountered another issue related to the quality of the scanned document images. Most of the documents were scanned a decade ago using outdated standards and methodology. In some cases, non-automated correction systems were used. This resulted in the scripts in OCR'ed documents being recognised less accurately. We began re-digitising the low-quality scanned images, a process requiring additional human and technical resources. The library digitisation team developed a list of necessary changes based on their evaluation of the quality of the scanned document images to recapture the documents and obtain better images.

1. Blur - blur may be caused by motion or the image being out of focus during capture.
2. Skew - if the document is not placed straight in the scanner, the scanned image may be skewed. It usually happens when the binding of a book or a bunch of newspapers is compressed too much.
3. Overexposure - overexposure produces an image that is too bright and has low contrast.
4. Aspect ratio - documents usually have a fixed aspect ratio. Sometimes, incorrect camera settings cause the aspect ratio to be stretched.

Calfa is a French company that provides AI and NLP solutions for oriental languages. They provide organisations, companies, and cultural heritage professionals with new technologies for non-Western languages. As a company that manages projects to preserve written and printed cultural heritage, the Calfa team has developed software that can automatically evaluate and improve the quality of damaged or poor-quality scanned images.

[2] https://github.com/portmind/armenian-ocr, Last visited: 5 December 2025

This innovative approach was developed in partnership with the TUMO Centre for Creative Technologies. TUMO is a free educational program that empowers teenagers to take ownership of their learning journey. The program comprises self-learning activities, workshops, and project labs that focus on 14 learning objectives. The TUMO workshop, 'Preserving the Past with AI', led by Chahan Vidal-Gorène and Baptiste Queuche, was carried out in collaboration with the National Library of Armenia[3].

Benefits for the National Library of Armenia from AI Partnerships

The National Library of Armenia stands to gain significantly from AI-powered digitisation and preservation projects. These collaborations offer technological improvements, institutional growth, increased international visibility, and stronger public engagement.

One of the most immediate benefits is improved digitisation quality. By integrating advanced, AI-driven tools, the library can achieve greater accuracy in optical character recognition, particularly for Armenian scripts, historical newspapers, and other materials. The result is more reliable digital collections and improved searchability for researchers and the general public. The integration of open-source AI tools further enhances this benefit by providing cost-effective solutions that can be adapted and expanded over time.

Another important outcome is capacity building and technical training. Library staff and collaborators will gain practical experience in using AI for digitisation and metadata enrichment. This will build internal expertise and establish a pool of skilled professionals capable of sustaining future innovation. Involving young specialists in these projects will also encourage engagement with cultural heritage among young people, ensuring the library remains visible and relevant to future generations.

The project will also raise the library's international profile. By collaborating with global institutions, research networks, and technology companies, Armenia will enhance its reputation as a nation that is committed to safeguarding its cultural heritage using cutting-edge methods and is forward-looking. By positioning itself as a partner in international projects, the National Library will access new opportunities for funding, expertise exchange and long-term cooperation.

Equally important will be the practical results. AI tools for digitisation and OCR will enable faster and more efficient workflows, allowing the library to digitise large collections that were previously out of reach. This will improve digital access to Armenia’s cultural heritage, making it available not only to scholars but also to the wider Armenian diaspora and global audiences. Ultimately, these benefits will strengthen the institution, creating a National Library with skilled staff and collaborators, enhanced international recognition, and greater visibility among young people and the general public. By strategically embracing AI partnerships, the library can ensure that Armenia’s written heritage is preserved and shared dynamically with the world.

[3] https://github.com/CVidalG/workshop-TUMO2025, Last visited: 5 December 2025

Collaborating with Portmind, Calfa and TUMO will integrate youth training and public–private synergy, offering a replicable model for underrepresented scripts in the digital sphere. Finally, we will explore the future scalability of the model and its policy-level implications for other low-resource cultural archives. By combining academic research, technical innovation, and pedagogical engagement, NLA projects offer a model that can be replicated by libraries worldwide.

Trond Myklebust:
The National Library of Norway as a Provider of Large Language Models

Artificial intelligence (AI) is being rapidly introduced across society, including in libraries. Many of the largest libraries are exploring how they can use AI to perform tasks that currently require significant human effort. This ranges from traditional cataloguing work to providing better services to library users. There is no doubt that libraries, like the rest of society, will change due to AI.

At the same time, AI requires digital content to train its models. In particular, large language models need vast amounts of digital textual content to achieve sufficient performance. These models are shaped by the content they are trained on, and to a very large extent, this content is harvested from the internet. Naturally, the largest languages and cultures will dominate such content, while smaller languages and cultures will have a correspondingly smaller influence on the models. These language models are not just about linguistic features but also about the culture, knowledge, and practices that the languages represent. Therefore, the language models reflect the societies from which the training data originates. The models will be used in many parts of society and will influence work methods, decisions, and cultural expressions. In some of the services based on large language models, such as ChatGPT, Deepseek, and Grok, clear biases can also be observed, which are undesirable in democratic societies.

The National Library of Norway (NB) is in a particularly favourable position to contribute to the development of language models that better reflect Norwegian languages and national conditions than the American- and Chinese-dominated models. Both the long-term digitisation program within the library and advanced digital legal deposit have contributed to a large, diverse, and up-to-date digital collection of published content. Almost all Norwegian literature is now available in digital form, and nearly 90% of all newspapers published in Norway are part of NB's digital collection. Combined with broad and deep knowledge of the digital collection, this is an extremely good starting point for training language models for Norwegian languages and Norwegian society. NB has through experiments and research shown that the content of the digital collection may serve very well as training data for large language models.

Therefore, the Norwegian government has asked NB to provide language models developed for use in Norway and for Norwegian conditions. By 2025, NB will establish a new unit whose main task is to train large language models. The models will be shared freely and made available for use by everyone. The Norwegian Parliament has allocated funds for both organisation and computing power. NB will combine this new task with internal personnel resources already dedicated for AI innovation and research. NB believes that the combination of research, innovation, and delivery to society will create a powerful unit. In total, the new AI unit will have approximately 20 employees.

NB will build up some internal computing capacity to support this work. However, training large language models requires significantly more computing power than NB can build on its own. Therefore, NB will use national and international infrastructure (EuroHPC) for research and innovation. The Norwegian Parliament has also allocated funding for such computing power.

Naturally, there are limitations on the use of content for training language models in Norway, as in other countries. NB, like others, must respect copyright, which largely restricts the legal use of much of the content in NB's collection. To understand the effect of using copyrighted content, NB, together with Norway's leading research environments, has conducted the Mimir[4] research project. A total of 17 models were trained on different parts of the content, and the effect on performance was compared across these models. The results are interesting and have been made openly available in reports and articles. It seems that it may make sense to tailor some models to use cases through building specific and crafted training sets.[5]

Based on the results from Mimir, the Norwegian government has asked NB to negotiate with Norwegian rights holders with the aim of purchasing the rights to train language models. The goal is for NB to be able to use content from recent literature and newspapers in the task described above, thereby training models that best fit Norwegian languages and conditions. The negotiations began in early 2025, based on principles agreed upon by NB and the rights holders following the Mimir project in 2024.

Finally, as mentioned earlier, large languages dominate the training of large language models. Languages and cultures with limited digital presence, particularly many minority languages, are falling behind in digital development. This could develop into a significant practical and democratic problem because AI is becoming a central component in many contexts. In Norway, we have several minorities, and the Norwegian Parliament has explicitly asked NB to work on language models for Sami languages. To address this challenge, NB is working with Sami-language communities to improve the quantity of training data for Sami languages and with research environments to find the best technological solutions for Sami languages.

[4] Mimir technical report: https://www.nb.no/content/uploads/2025/12/The-Mimir-Project-Technical-Report-Short-eng-.pdf, Last visited: 8 December 2025

[5] The Impact of Copyrighted Material on Large Language Models: A Norwegian Perspective: https://arxiv.org/abs/2412.09460, Last visited: 5 December 2025

Taner Beyoğlu:
The Leading Role of the National Library of Türkiye in the Age of Artificial Intelligence

Introduction

The twenty-first century is witnessing a profound transformation in the way libraries define their mission, deliver their services, and interact with society. No longer perceived merely as passive repositories of books, libraries are increasingly recognised as living institutions that adapt dynamically to technological, social, and cultural changes. Within this evolving landscape, the National Library of Türkiye (NLT) has positioned itself as a pioneering actor. By strategically embracing artificial intelligence (AI), the NLT not only modernises its internal operations but also provides a replicable model for more than 1,300 public libraries under the Directorate-General for Libraries and Publications. This dual responsibility situates the NLT at the heart of national and international debates on how AI can serve as a transformative force for libraries in the digital era.

Institutional Framework and Vision

The NLT operates within the Directorate-General for Libraries and Publications, which is the central authority responsible for library policy across Türkiye. This governance structure allows the National Library to function as both a flagship institution and an incubator for innovation, piloting new services that can later be disseminated nationwide. The library itself occupies a total area of 39,000 square meters, of which 9,000 are open to the public, and offers seating for more than 1,300 users. Yet the physical presence of the institution is only one dimension of its mission. Its primary objectives include supporting national cultural research, facilitating scientific and artistic studies, ensuring the flow of information domestically and internationally, and contributing to the formation of cultural policy.

The true ambition of the NLT extends beyond the preservation of heritage to the pursuit of equity in access. In this vision, digital transformation plays a central role, ensuring that knowledge reaches every citizen regardless of geographical location, social status, or physical ability. The library thus embodies the dual identity of being both a guardian of the past and a laboratory for the future.

Building the Digital Twin of the National Library

A cornerstone of this transformation is the National Digital Library System, launched in 2014. Its guiding principle is the creation of a digital twin of the National Library of Türkiye, enabling remote access to collections of national significance. Over the past decade, this initiative has digitised 50,000 volumes of books, 200,000 manuscripts, two million periodicals, and 70,000 non-book materials. Among its achievements are the digitisation of more than 90,000 Ottoman Turkish works, thousands of historical negatives, volumes of rare periodicals, and nearly 5,000 gramophone records.

With more than one hundred digital access stations distributed across the provinces and located within public libraries, the project reduces inequalities of access and ensures that cultural resources are available in every corner of the country. By combining advanced digitisation technologies with nationwide outreach, the NLT has laid the groundwork upon which AI-supported services could later be developed.

The Role of AI: Innovation and Incubation

In the age of AI, the NLT assumes a dual role. First, it integrates artificial intelligence into library services in a manner that is both ethical and effective, enhancing cataloguing, collection development, and user experience. Second, it functions as a national incubation centre where emerging tools are tested, refined, and scaled. This ensures that innovations do not remain confined to the national library itself but become resources for the wider network of public libraries.

By situating itself at this intersection of innovation and dissemination, the NLT effectively bridges the gap between global technological trends and local community needs. Its strategy demonstrates that AI, when carefully implemented, can reinforce rather than replace the human values that lie at the heart of librarianship.

The 100 AI-Based Smart Libraries Project

Perhaps the most ambitious initiative in this field is the "100 AI-Based Smart Libraries Project," launched in 2024. Designed to embed artificial intelligence across a wide spectrum of library functions, the project equips one hundred public libraries throughout Türkiye with innovative tools. These include AI-supported kiosks for catalogue search, borrowing, and returning; the digital library assistant "Kitabi"; and AI-driven processes for collection development, technical services, and user feedback analysis.

Developed in collaboration with Microsoft Azure and Cerebrum Tech, the project unfolds in three phases, focusing sequentially on user experience, technical services, and intelligent interfaces. Its significance lies not only in technological innovation but also in the democratisation of knowledge, as it provides user-centered, responsive, and accessible services across diverse local contexts.

Expanding the Smart Library Concept

The smart library vision extends beyond AI. It encompasses RFID-based tracking systems, inclusive indoor navigation tools for individuals with disabilities, smart book platforms, national digital library stations, and controlled access systems. Additional innovations such as online reservation platforms, IP telephony integration, and even library robots exemplify the multidimensional nature of this transformation. These technologies collectively enhance personalisation, efficiency, and inclusivity, reshaping the library experience for contemporary users.

Commitment to Accessibility

Central to the NLT's philosophy is the conviction that innovation must promote inclusivity rather than exclusion. This is reflected in services such as the "Talking Library," which provides over 6,000 audiobooks for visually impaired users; "Kitap+," a sign-language platform with 100 animated works for hearing-impaired communities; and the "E-Kitabım" mobile application, offering access to more than 5,000 e-books. Moreover, the library curates digital exhibitions featuring multilingual content and internationally renowned artworks, opening cultural heritage to global audiences. These initiatives testify to the library's role in democratising culture and knowledge through technological means.

AI Services in Action

Artificial intelligence is no longer a future aspiration but a present reality in Türkiye's library ecosystem. Chatbot Kitabi, integrated kiosks, and AI-supported technical processing are already operational, assisting both users and staff in navigating vast collections and complex workflows. These tools not only improve efficiency but also generate data that can inform decision-making, enabling libraries to anticipate user needs and adapt their services accordingly.

Broader Implications and Future Directions

The transformation of the National Library of Türkiye is designed as a scalable and shareable model, offering inspiration for both national and international contexts. By combining technological innovation with institutional leadership, the NLT has developed a blueprint that can be adopted by libraries in underserved and remote regions. Future priorities include the further expansion of AI-based services, the improvement of data interoperability, and the development of user-centred digital tools that align with both national needs and international standards.

Libraries as Living Systems

At the core of these efforts is a vision of libraries as living systems. They evolve in response to societal transformations, technological disruptions, and cultural shifts, yet they remain anchored in timeless values of accessibility, knowledge sharing, and inclusivity. The integration of AI into library services does not diminish these values; on the contrary, it reinforces them by offering new ways to connect knowledge with people.

Conclusion

By embracing AI as a transformative force, the National Library of Türkiye demonstrates that the future of libraries lies not in choosing between tradition and innovation, but in synthesising both. It

is possible to preserve cultural heritage while simultaneously building intelligent, inclusive, and interconnected services that reflect global technological progress. The experience of the NLT illustrates how national libraries can act as key hubs for intelligent solutions, ensuring that knowledge—humanity's most valuable asset—remains universally accessible in the digital age.

Yurii Kovtaniuk:
Application of Artificial Intelligence in the Activities of V. I. Vernadskyi National Library of Ukraine

In view of the rapid development of Artificial Intelligence (AI) technologies worldwide and the growing interest of the global library community in the application of AI in library activities, in particular in European national libraries and in the Conference of European National Librarians (CENL) as a consolidating body of European libraries, over the past two years V. I. Vernadskyi National Library of Ukraine (VNLU) has taken certain steps to study this issue, identifying VNLU's strengths and weaknesses, and analysing existing opportunities and potential threats. In a distant approximation, this resembled the SWOT analysis introduced for discussion by colleagues from the University of Sheffield in 2023[1]. The first extensive report on this research was presented at VNLU International Conference in October 2024[2]. A large presentation was prepared for the report, a substantial part of which was devoted to existing tools for automating the activities of VNLU, which could serve as a foundation for the development of AI. It became clear that the issue required more detailed and comprehensive study, given that at that time no thorough research on the application of AI in library activities had been conducted in Ukraine, apart from a few studies focusing on specific aspects of AI.

Separately, the issue of applying AI in the activities of archives, libraries, and museums was discussed at the meeting of the Academic Council of VNLU in February 2025.

The next steps focused on studying the legislative acts of Ukraine concerning the development of AI as a legal basis for promoting these technologies in the library field. Special attention was given to scientific research conducted by specialised institutions, particularly their proposals for the overall strategy for the further development of AI in Ukraine, including the Strategy for the Development of Artificial Intelligence in Ukraine, published by the Institute of Artificial Intelligence Problems[3]. Unfortunately, no mention of librarianship was found in the documents examined.

To obtain the results of the comparative analysis in 2025, the legislative acts of developed countries were reviewed, along with the strategic visions for the implementation of AI provided by the United Nations (in particular UNESCO), the Council of Europe, the African Union, the U.S. government agency National Archives and Records Administration (NARA), the International Federation of Library Associations and Institutions (IFLA), the Conference of European National Librarians (CENL), the International Council on Archives (ICA), national and academic libraries of many countries, as well as the international initiative Artificial Intelligence for Libraries, Archives and Museums (AI4LAM).

---

[1] https://orda.shef.ac.uk/ndownloader/files/43905486, Last visited: 9 December 2025

[2] Kovtaniuk Yu. S., Kuznietsov O. Yu. (2024). Vykorystannia shtuchnoho intelektu v diialnosti arkhiviv, bibliotek i muzeiv. Biblioteka. Nauka. Komunikatsiia. Intehratsiia u mizhnarodnyi bibliotechnyi prostir: Tom 1 : materialy Mizhnar. nauk. konf. (Kyiv, 8–10 zhovt. 2024 r.) : u 2 t. / NAN Ukrainy, Nats. b-ka Ukrainy im. V. I. Vernadskoho, Asots. b-k Ukrainy ; vidp. red. O. M. Vasylenko, vidp. sekr. M. V. Ivanova. Kyiv, S. 281–286. URL: https://irbis-nbuv.gov.ua/everlib/item/er-0004933, Last visited: 5 December 2025

[3] (2023). Stratehiia rozvytku shtuchnoho intelektu v Ukraini: monohrafiia [A. I. Shevchenko ta in.] Kyiv: Instytut problem shtuchnoho intelektu, 305

The results of this study were published in the form of a scientific article in the second half of 2025[4]. The authors of this work state that almost all intergovernmental unions (EU) and international organisations (UN, UNESCO), including sectoral organisations (IFLA, ICA), regional organisations (CENL), and intersectional initiatives (AI4LAM), have established special groups for the development of AI in specific fields or in archives and libraries more broadly. Active discussions are underway on the possibility of formulating conceptual foundations for the further development of AI, disseminating practices of AI implementation in archives and libraries, sharing experiences, identifying the most successful solutions and ways to improve the results achieved, and drafting international recommendations. Many of the world's leading archives and libraries have developed their own AI strategies and have begun gradually implementing them in their operations.

In order to obtain funding and conduct fundamental research on the problem of applying AI in library work in Ukraine, VNLU prepared and submitted to the National Academy of Sciences of Ukraine (NAS of Ukraine) a request for the implementation of scientific (scientific and technical) work on the topic "Theoretical aspects and applied principles of using artificial intelligence in library work" in 2026–2028. The proposed topic was approved by the decision of the Bureau of the Department of History, Philosophy and Law of the NAS of Ukraine.

The request outlined the overall contributions of European and Ukrainian librarians and scholars, defined the plan for the implementation of the topic and its objectives, as well as the expected results concerning the application of AI both in the activities of VNLU and in other institutions, not limited to libraries. In addition to theoretical outcomes, based on the existing capacities of VNLU, practical tasks were planned to be addressed, including the creation of specific software tools that would be freely accessible to users online through the Library's web portal.

Considering that VNLU is simultaneously the largest universal library in Ukraine and the largest research institution within the system of the National Academy of Sciences of Ukraine (NAS of Ukraine), this has enabled the Library to create the country's largest scientific knowledge resources, comprising over five million electronic records, about two million of which are full-text. These are primarily periodicals and monographs that are freely available to users online through VNLU web portal. Today, the resources of VNLU provide up-to-date information for most scientific research conducted in Ukraine across almost all fields of knowledge, as well as for virtually all educational programs in Ukrainian secondary and higher education institutions, including postgraduate studies. The established knowledge base serves as a solid informational foundation for the development and implementation of AI technologies in the activities of VNLU.

One of the routine tasks in preparing scientific texts is the compilation of classification indices based on the Universal Decimal Classification (UDC), which, in accordance with the National Standard of Ukraine DSTU 4861:2007 "Information and Documentation. Publications. Imprint Data", must be included in the storage code of a publication and appear in the imprint of every edition. Therefore, all scientific texts published in professional journals in Ukraine are required to include a UDC index.

[4] Dubrovina, L. A., Kovtaniuk, Yu. S., & Harahulia, S. S. (2025). Regulatory and Legal Framework, Research, and Prospects for Artificial Intelligence Implementation in Archival and Library Affairs in Ukraine: an Analytical Review. *Sci. innov.*, 21(5), 18–32. https://doi.org/10.15407/scine21.05.018, Last visited: 5 December 2025

This is a complex task that requires the efforts of specialists working with special tables[5]. When authors compile the UDC index independently, the level of inaccuracies is high. The use of such UDC indices leads to incorrect identification of texts in the future. In 2024–2025, staff members of the VNLU developed software that implements a clustering method (identifying groups of identical text fragments by content) for checking the UDC indices of publications, which enables, with high probability and without expert involvement, the determination of the accuracy of the correspondence between the UDC index and the stated text[6]. An important factor is that the check is carried out not by the annotation, which briefly indicates the content of the scientific work, but by the entire scientific text, the volume of which affects only the time of its processing. It should be noted that most experts assign UDCs by the annotation. Therefore, the proposed method is much more accurate. In addition, this method can automatically obtain an assessment of the completeness of the disclosure of the declared name in the scientific text.

The proposed UDC checking tool can serve as the basis for creating an online service for assigning UDC indices to any text submitted by its author for verification. At a minimum, it will be applied in the publications of VNLU.

The application of similar methods also appears promising for improving the indexing of various information objects and specific data within separate structures of knowledge resources. This represents a significant step toward the development of intelligent search systems that can help modern librarians move closer to solving the problem of underutilised collections, given the complex procedures of providing access through analogue technologies and the worsening situation caused by users' growing expectations for expanded digital content. This occurs against the background of a reluctance to access information in traditional ways, reflecting the decline of reading culture among younger generations of users. A similar stagnation can be observed in access to digital information. As the volume of digital content increases without the application of advanced search systems capable of providing semantically relevant results to user queries, the process begins to resemble reading analogue originals, with the only difference being that the reading takes place on a screen. This is particularly true for non-textual information.

There is a strong conviction that the next step in mastering the growing volumes of information is the improvement of search systems. Methods of linear search in full-text arrays, even with the gradual increase in computer performance, are becoming increasingly ineffective. Although indexing is a good alternative for big data, it is not a panacea for solving the problem of intelligent search, as it excludes the semantic component from the process of information retrieval. However, it should be understood that these methods cannot be abandoned at the present stage, since they provide the informational basis for accumulating data on the semantics of queries and the system's responses to different requests, taking this semantic component into account. Thus, the focus is on the

[5] https://udcsummary.info, Last visited: 5 December 2025

[6] Kuznietsov O., Zaika V. Vyznachennia indeksiv UDK novykh nadkhodzhen v elektronnomu vyhliadi dlia formuvannia elektronnoi biblioteky prohramnymy zasobamy. Bibliotechnyi visnyk. 2023. № 3. S. 3-16. doi: https://doi.org/10.15407/bv2023.03.003, Last visited: 5 December 2025; Zaika V., Kuznietsov O., Handziuk V. Systematyzatsiia povnotekstovykh nadkhodzhen do elektronnoi biblioteky metodamy klasteryzatsii. Vidkryta nauka ta innovatsii. 2024. T. 1. № 2. S. 4–12. http://doi.org/10.62405/osi.2024.02.01, Last visited: 5 December 2025

accumulation of paradata and the improvement of search tools, considering the results of processing such data. As more paradata are accumulated, the search becomes higher in quality, semantically adequate to the query, and more intelligent.

A sufficient array of paradata enables the library to achieve full control over the user experience. The functioning of automated library systems will increasingly resemble that of popular global platforms such as Facebook, TikTok, YouTube, etc. In these systems, decisions about providing users with new content are made automatically by AI, and since these decisions are based on the analysis of previously accumulated information about user preferences and viewed content, most users tend to accept the content offered to them.

For scientists, access to intelligent information search tools means not only expanding the boundaries of the search, which ensures the completeness of the obtained search results on the principle of "not missing anything in the array being studied", but also on the basis of real multifacetedness, the development of which can be provided by paradata and modern software and hardware, making this search truly aimed at achieving a specific research goal. In fact, the more AI learns, the better the search robot will penetrate so deeply, without missing anything from the focus determined for it, the better the task will be formulated for it. Therefore, despite all the concerns about AI, human participation remains extremely important. And it is clear that non-standard human thinking will put AI in deadlock situations for a long time.

In addition, in the future, the development of these clustering methods may allow for the creation of its own plagiarism checking system, given the large array of double-text scientific works that are currently available in VNLU repository, the number of which is constantly increasing. However, this proposal is debatable, given the costs required to develop a new system. Currently, VNLU uses the StrikePlagiarism software solution[7]. In 2025, the range of services provided by StrikePlagiarism VNLU was expanded to include checking for the search of text generated by AI. Since the beginning of the year, the AI module has been tested and has proven highly relevant results in practice, which, according to experts, has already allowed us to achieve a high level of integrity of the authors of scientific texts of VNLU. Therefore, this module is now mandatory for checking all scientific texts published by VNLU.

Doubts about creating new software tools are relevant as long as alternative proposals for building new services exist. The main criteria will be the cost of the service, its quality, and the level of technical support. The same applies to AI: should new services be built on accessible general platforms or developed as specialised ones? In our view, each service requires a detailed separate discussion, which may lead to different approaches and solutions. No opportunities should be categorically rejected. The experience of the world's leading libraries demonstrates the successful integration of multiple solutions within a single information system.

At first glance, the service of generating bibliographic references to a publication in various formats of bibliographic description appears simple, yet it is in demand. The task can be solved within the framework of the transition to common tools for managing the processes of preparing periodicals

[7] https://strikeplagiarism.com, Last visited: 5 December 2025

and monographs with the subsequent formation of an archive of these publications to provide free access to them online. Such common open-source software solutions are Open Journal System (OJS) and Open Monograph Press (OMP), which provide the ability to automatically form a bibliographic description of publications in more than ten different formats. This is a useful tool for automating the preparation of scientific texts in the mode of copying a ready-made reference to a general list of references in a specified format. However, if there are a lot of references and they urgently need to be converted into another format of bibliographic description in accordance with the requirements of another publication, then this becomes a routine and an unpleasant addition to the creative process. Our task is to create an online service that allows scientists to translate ready-made lists of references from one format to another. Similar solutions are offered in Ukraine by various resources, for example https://vak.in.ua[8].

The previous task becomes more complicated if the researcher has a mixed set of references, some of which must be transliterated, for example, when preparing material for a repository to obtain a DOI. The majority of transliterated bibliographic references are currently formatted in accordance with the Resolution of the Cabinet of Ministers of Ukraine dated 27.01.2010 No. 55 "On the Systematisation of Transliteration of the Ukrainian Alphabet into the Latin Alphabet".

However, linguists rightly consider this method of transliterating Ukrainian texts to be simplified, which was created mainly for the issuance of international passports in Ukraine. This method does not fully convey the entire phonetic range of the Ukrainian language. If there is a need for an accurate representation of Ukrainian texts in the Latin alphabet, primarily scientific, especially those related to the study of ancient texts, it is here that researchers really face the problem of the lack of tools for transmitting the wealth of this information to a foreign-language information environment. In 2021, the national standard of Ukraine DSTU 9112:2021 was issued. Cyrillic-Latin transliteration and Latin-Cyrillic re-transliteration of Ukrainian texts. Writing rules that methodologically solve the specified problem and from 12 transliterations and re-transliteration tables that cover all periods of the existence of Ukrainian writing: Old Ukrainian (Old Church Slavonic), Old Ukrainian, Modern Ukrainian alphabets. The standard separately defines the transliteration of names and offers two systems - with diacritical marks (A) and with the use of letter combinations (B). System A is recommended for printed products (books, magazines, booklets, almanacs, directories, maps, atlases, etc.); website content; names of vehicles and other objects belonging to Ukraine; information systems; letterheads of legal entities; signs; historical monuments; postage stamps; addresses on postage. System A is recommended for use in e-mail; website and web page addresses; computer program scripts; identity documents; multilingual text corpora; bibliographic records; road signs; names of cities at airports[9]. However, to apply this standard, appropriate tools are required, the existence of which is not known to us.

[8] Last visited: 5 December 2025

[9] Vakulenko M. O. Vykorystannia natsionalnoho transliteratsiinoho standartu DSTU 9112:2021 v informatsiinomu zabezpechenni bibliotechnoi roboty. Biblioteka. Nauka. Komunikatsiia. Innovatsiini transformatsii resursiv i posluh : materialy Mizhnar. nauk. konf. (4–6 zhovt. 2022 r.) / NAN Ukrainy, Nats. b-ka Ukrainy im. V. I. Vernadskoho, Asots. b-k Ukrainy, vidp. red. O. M. Vasylenko. Kyiv, 2022. S. 12–15. URL: https://irbis-nbuv.gov.ua/E_LIB/PDF/er-0004349.pdf, Last visited: 5 December 2025

Unfortunately, the use of AI in cataloguing in Ukrainian libraries is in a state of flux, requiring in-depth study rather than understanding, requiring in-depth study rather than mere understanding, as expressed by Dr A. Cohen, Chair of the RDA/RSC Joint Working Group on AI: "AI is being used in cataloguing departments around the world, and we are grateful to the extended membership for their feedback on our initial survey. This White Paper is the first phase of our research into AI in this area, and we are pleased that both the RDA Council and the RSC have endorsed our continued work"[10]. For example, it is worth mentioning a question that was addressed to Yuriy Lukianchuk, Associate Professor at the Department of Computer Science of Lutsk National Technical University, who held a seminar on "The Application of Artificial Intelligence in the Professional Activities of Librarians", organised by the National Scientific Agricultural Library of the National Academy of Agrarian Sciences of Ukraine in February 2025[11]: Is it possible to use AI to support cataloguing in librarianship? The response was that this is currently not possible. From our side, no questions were raised to the information technology specialist. However, the direction of development of the Ukrainian library community is clear. In our opinion, on this path the support of CENL and other leading organisations in maintaining expert dialogue with the Ukrainian professional community, in particular through VNLU, as well as the possibility of engaging in open and constructive dialogue, is of great importance. Therefore, an important task is the introduction of AI into the cataloguing processes of VNLU and the dissemination of this experience to the activities of other libraries in Ukraine.

It should be noted that not the least role in this will be played by the employee of VNLU, Dr N. Strishenets, who is a member of the European RDA Interest Group (EURIG) from Ukraine[12]. In Ukraine, dialogue on the implementation of RDA is ongoing. The latest initiative was launched under the auspices of the Yaroslav Mudryi National Library of Ukraine to establish a working group on the implementation of the international cataloguing rules "Resource Description and Access (RDA) in Ukrainian Libraries", in which VNLU employees, including N. Strishenets, participated. It is expected that this will contribute to the wider implementation of RDA and enable the VNLU to integrate RDA with AI.

An important step towards enhancing the accessibility of textual information on web resources is the implementation of text-to-speech conversion, the provision of subtitles for video content, and the use of automatic translation systems, including the transformation of textual information into audio and vice versa. Gradually, resources of the Ukrainian information space are being equipped with such features, supported by the state leadership. A large-scale task is the modernisation of the official web portal of the VNLU, within which this objective can be realised.

Among the plans is also the introduction of communication with users through a chatbot. Which AI system will serve as the foundation, ChatGPT or another, remains a subject for future discussion.

---

[10] (2025). New White Paper: Considering AI in the Growth of RDA. ALA, CFLA, and CILIP: Chartered Institute of Library and Information Professionals. Retrieved from https://www.rdatoolkit.org/rsc/papers2025_1, Last visited: 5 December 2025

[11] https://dnsgb.com.ua/novini/podii/2025/shtuchnij-intelekt/, Last visited: 5 December 2025

[12] https://www.rdatoolkit.org/europe/members, Last visited: 9 December 2025

Therefore, an important task at the current stage of the VNLU's development is the interdisciplinary interaction between the developers of the above-mentioned tools and librarians, archivists, and humanities researchers, who should confidently transfer routine work to automated means. This makes it possible to free up time for the realisation of their creative potential, to significantly expand the horizons of their own research, and, for the more ambitious, to propose larger-scale projects. In order to apply the powerful computational capacity of modern computers, software, and AI in their research, humanities scholars must, undoubtedly, broaden the scope of their concepts in related fields of knowledge, primarily in computer science. The principal aim of such interdisciplinary engagement is to acquire the ability to formulate new tasks for the developers of automated systems, software, and neural networks.

The application of new technologies in the humanities is an objective process, as is the integration of knowledge and the consolidation of efforts to create relevant knowledge resources. What is subjective, however, is the management of these processes. The quality of such management depends on the readiness of specialists from different fields of knowledge for dialogue, their mutual interdisciplinary engagement, and their willingness to understand ideas arising in other areas of activity, all of which contribute to achieving a common result.

In our view, humanities scholars should become a driving force in such interdisciplinary interaction, given that the discussion concerns the development of digital humanities, the creation of new research tools specifically in the humanities domain, and the introduction of advanced services for library users. We consider that an overbearing attitude of specialists from a particular field of knowledge can be a detrimental factor in such dialogue. It should be emphasised that, just as humanities scholars must overcome barriers to gaining sufficient competence in information technology, so too should information technology specialists restrain themselves from a tendency to oversimplify and formalise problems, which often arises from an insufficient depth of engagement with the humanities. Experience demonstrates that true breakthroughs are achieved in teams where an atmosphere of mutual understanding among specialists from different fields of knowledge prevails.

Hannes Lowagie:
AI Integration in Legal Deposit Workflows: Enhancing Cataloguing Through Automation and Collaboration

Introduction

National libraries face a rising tide of information. Every year, more different materials are deposited, in more languages and in more formats. Traditional cataloguing workflows, while thorough and well-established, are increasingly under pressure to scale, adapt, and standardise. At the Royal Library of Belgium (KBR), we have turned to artificial intelligence not just to cope with this growth but to reimagine how legal deposit and cataloguing workflows can function in a scalable, collaborative future.

This article outlines our approach to integrating AI into legal deposit workflows, highlighting three core applications of AI and the benefits we have observed in practice. Our aim is not only to modernise cataloguing but also to foster shared standards, collaboration, and interoperability across national libraries.

Why Modernise Cataloguing Workflows?

Legal deposit has been a cornerstone of national bibliographic services for decades. At KBR, since 1966, authors and publishers have been required to deposit their works. This has created a rich and growing collection, but different challenges arise like a growing volume of incoming publications, an increasing diversity of languages and formats, and a need for standardised and scalable workflows to ensure consistency and quality across records.

The goal of our modernisation effort is to combine human expertise with AI's ability to automate repetitive tasks, streamline metadata processing, and enhance classification and discoverability.

Overview: Three Ways AI Assists Human Cataloguers

Our AI-supported workflow focuses on three main areas:

1. Metadata recognition for fast registration
2. Metadata improvement aligned with internal cataloguing rules and RDA (Resource Description and Access)
3. Automatic classification to improve discoverability and ensure consistency

Each function has been carefully designed to support human cataloguers, not to replace them. The aim is to enhance the cataloguing process by combining the efficiency and speed of artificial intelligence with the critical thinking and expertise of trained professionals. AI takes on the more

laborious and repetitive tasks, such as recognising metadata, applying formatting rules, or suggesting classifications, freeing up valuable time for human cataloguers to focus on the tasks that require nuanced judgement, contextual understanding, and subject knowledge. Ultimately, the final responsibility for validation and bibliographic accuracy remains firmly in human hands. This ensures that while workflows become faster and more consistent, the quality and integrity of the bibliographic records are maintained to a high standard. The partnership between human and machine is not a compromise, it's a way to get the best of both worlds.

AI 1: Metadata Recognition for Fast Registration

One of our first priorities was accelerating the initial metadata capture. To do this, we developed a custom document processing model using a low-code Microsoft platform: Power Apps and Power Automate, which integrate easily with AI components.

Key features:

- Trained on local data: Using a minimal dataset of as few as 5 examples, a model can be learned to detect metadata from images of documents (e.g., title pages). While 5 is a bare minimum, the variety of title pages has learned us that a larger number is needed (approximately 250-300 images).
- Dynamic rule creation: The AI does not rely on static templates. Instead, it creates rules based on document layout, text placement and font size, but also semantic content (edition statements were recognised based on the word 'edition')

This document processing approach can significantly accelerate the initial phase of cataloguing, especially when it comes to extracting key metadata elements. By automating these early steps, we save valuable time, improve consistency, and ensure that important metadata is captured efficiently. After the detection the detected metadata (e.g., author, ISBN, place names) can be used in HTTP queries to enrich records via external databases such as other bibliographic catalogues of national libraries, ISNI and GeoNames.

However, the pace of progress in artificial intelligence is remarkably rapid. The recent emergence of advanced models such as GPT-4o, along with increasingly powerful multimodal tools that can process text, images, and other formats simultaneously, opens up even more exciting possibilities. These technologies have the potential to greatly expand our capabilities in metadata detection and enrichment, for example, by interpreting scanned title pages more intelligently, recognising layout and design cues, or even understanding handwritten or stylised text. As these tools evolve, they could transform not only how metadata is captured but also how it is contextualised, interpreted, and connected across systems, bringing us closer to a future where cataloguing becomes smarter, faster, and more collaborative than ever before.

AI 2: Improving Metadata to Follow Cataloguing Rules

The second application of our AI integration focuses on refining and correcting metadata in line with both our internal cataloguing guidelines and the RDA (Resource Description and Access) Application Profile. This step is crucial for ensuring that metadata not only meets international standards but also aligns with the specific requirements and practices of our institution.

At the heart of this process is a structured JSON file that encodes our cataloguing rules in a machine-readable format. For each RDA element, the JSON includes key attributes such as, recording methods, obligation status (whether the field is mandatory or optional), repeatability value encoding schemes (VES), and other metadata-specific requirements. In addition to these technical specifications, the JSON also incorporates our internal cataloguing policies, which guide how certain fields should be interpreted or formatted according to local practices. This JSON file functions as the backbone of our entire cataloguing framework, serving a dual purpose. On the one hand, it supports a human-readable HTML interface that retrieves values from the JSON and presents them in a clear, accessible format. This serves as our primary cataloguing guide, where cataloguers can easily consult up-to-date documentation and instructions while working on records. On the other hand, the same JSON is also directly integrated into our AI-powered cataloguing workflows. The rules it contains are used to construct prompts for large language models (LLMs), effectively turning our internal documentation into dynamic, machine-readable instructions. This ensures that both human cataloguers and AI tools are working from the same source of truth, applying identical rules and standards across the board.

By unifying the cataloguing process in this way, we not only enhance consistency and transparency but also pave the way for more efficient collaboration between people and machines. This approach helps bridge the gap between traditional expertise and automated processing, creating a smarter and more adaptable workflow that respects both human judgment and the power of AI.

An added advantage of using a structured JSON format to encode cataloguing rules is that it allows for broader interoperability and reuse across different institutions. For instance, consider an application capable of recognising metadata with high precision and transforming it according to a set of rules defined in a JSON file. If one institution has already implemented such a tool using its own cataloguing guidelines encoded in JSON, another library or organisation could adopt the same tool and simply swap in their own JSON file to match their specific application profile and internal cataloguing policies. This means that the core application and methodology remain the same, while the behaviour of the tool is tailored through the configuration file. This approach promotes efficiency, flexibility, and collaboration, making it easier for libraries to adopt shared technologies without sacrificing their unique workflows and standards. It opens up possibilities for shared development, standardised infrastructure, and a more scalable and sustainable approach to cataloguing in the age of AI. It is not only about sharing standardised data, but also about sharing the cataloguing rules that explain how that data was interpreted, modified, and recorded. Providing access to the underlying rules and guidelines adds an essential layer of transparency and context—allowing others to understand not just what the data contains, but also how and why certain decisions were made during its creation. This kind of openness supports interoperability across institutions, especially

when working with shared or aggregated datasets. Making both the data and the decision-making frameworks behind it available, can foster a more coherent and efficient bibliographic ecosystem.

AI 3: Automatic Classification for Discoverability

The third major application of AI is automated subject classification. This is especially important for maintaining the Belgian Bibliography and improving users' ability to discover relevant materials. Our model processes imported or detected metadata, then applies a local classification scheme. Our classifications are maintained in a structured CSV file, meaning the output of any automated process must correspond to one of our local classification codes, not just a random or generic suggestion generated by an AI prompt. This ensures that classifications remain consistent with our established taxonomy and are usable within our existing cataloguing systems. One of the key benefits of using a prompt-based approach within an automation tool like Microsoft Power Automate is the ability to inject variables directly into the prompt. This makes the workflow highly customisable and context-aware. In our case, the variable used contains a CSV file with two important columns: a classification identifier (which maps to our internal code) and an explanation (which describes the meaning or scope of the classification).

In the prompt, we explicitly instruct the model to consider the explanation when selecting the most appropriate classification. This ensures that the model isn't just matching keywords, but is actually engaging with the intent and context behind each classification category. By doing so, we retain a high level of control over the output, and the results become more tailored to our specific cataloguing logic. Moreover, this method gives us a practical way to improve classification accuracy over time. By expanding or refining the explanatory notes in the CSV file, we help the model better understand how we interpret certain terms or concepts. This, in turn, leads to more accurate, consistent, and institutionally relevant classifications, even as the range of content types and topics continues to evolve.

This method ensures that all records are classified using a unified and consistent system. By applying the same classification structure across the entire catalogue, we maintain coherence and standardisation, which is essential for both internal workflows and public-facing search tools. Importantly, the system is designed to be scalable, meaning that even when processing large volumes of data, the classification remains reliable and consistent. At the same time, there is always room for human intervention. A professional cataloguer can review, adjust, and refine the automated classifications whenever necessary, adding a critical layer of quality control and domain expertise to the process. This hybrid approach, combining automation with human oversight, not only streamlines the workflow but also enhances the discoverability of resources across various platforms and interfaces. Users benefit from more accurate search results, better organisation of materials, and improved access to content, regardless of how or where they are interacting with the library’s collections.

Supporting Tools: Microsoft Power Platform

A key enabler of this entire approach has been the Microsoft Power Platform, which provides a flexible, low-code environment that supports rapid development and integration of intelligent tools. It brings together several core components that play a vital role in streamlining and enhancing our workflows:

- Power Apps: This tool allows us to build custom applications tailored to our specific cataloguing needs, all without requiring full-scale software development. With its intuitive drag-and-drop interface, staff members can create user-friendly apps that simplify complex tasks.
- Power Automate: Serving as the orchestrator of workflows, Power Automate enables seamless integration between different systems and data sources. It can trigger AI models, manage task sequences, and respond to user or system events—automating repetitive processes while ensuring reliability and efficiency.
- AI Hub: This feature provides access to a wide range of prebuilt AI models, but more importantly, it supports the development and deployment of custom models. For example, we have created document processing models specifically trained to extract metadata from title pages and copyright pages. More recently, they have integrated GPT technology, allowing us to embed advanced language models into automated workflows. This empowers intelligent decision-making by enabling AI to understand and act on nuanced textual instructions.

Together, these tools create a robust and adaptable infrastructure that enables even non-developers to contribute meaningfully. This is especially valuable for so-called "Citizen Developers" (staff who possess deep domain knowledge but may not have traditional coding skills). With the Power Platform, they can actively build, customise, and refine solutions, ensuring that the tools we develop remain closely aligned with real-world library needs and evolving cataloguing practices.

Benefits Achieved

Since integrating AI into our cataloguing workflows, we have observed a range of tangible and measurable improvements that have enhanced both efficiency and quality. One of the most notable benefits has been a marked increase in consistency. Bibliographic records can now follow a more uniform input-structure and format, which significantly reduces the risk of errors, inconsistencies, and duplication across the catalogue. Another key improvement is scalability. The AI-enhanced workflows are capable of handling much larger volumes of material without requiring a proportional increase in manual effort. This is particularly important given the steady growth in the number and variety of publications we receive through legal deposit. Tasks that were previously time-consuming and labour-intensive can now be performed much more efficiently, freeing up human resources for more complex decision-making and quality assurance. It can also help achieving greater alignment with international metadata standards like RDA or ISNI. Perhaps most exciting is the potential for broader collaboration. The AI models, training datasets, and cataloguing rules we develop can be

shared with other institutions, enabling the co-creation of tools and methodologies. This opens up valuable opportunities for joint development, innovation, and standardisation within the wider library and information community. By sharing resources and aligning efforts, we can collectively advance the field and build more sustainable and intelligent cataloguing systems.

Conclusion

AI integration in legal deposit workflows is not about replacing human cataloguers. It's about augmenting human judgment with intelligent automation, ensuring that even as our collections grow, the quality and discoverability of metadata remains high.

At KBR, our experience shows that with the right tools and mind set, AI can serve as a valuable collaborator and a catalyst for deeper cooperation across institutions. Let’s work together to build the future of library cataloguing.

More about our application and method can be found in my recently published book “AI-Powered Cataloguing” [1].

[1] AI-Powered Cataloguing: https://www.facetpublishing.co.uk/page/detail/ai-powered-cataloguing/?k=9781783308071, Last visited: 5 December 2025

Kalina Ivanova and INSAIT: Collaborative AI

The National Library of Bulgaria „St. St. Cyril and Methodius" is the oldest cultural institute of post-liberation Bulgaria, established in 1878. We are proud to be the largest public library in the country. We are well aware of the challenges ahead, but we also take pride in our achievements and ongoing development. Our library sees itself as a centre of information and research, dedicated to serving the public, preserving national memory, and promoting education and culture in all their forms. With nearly 9 million items, our collection represents the largest book treasury in Bulgaria. Of particular significance is our Oriental Collections Department, which houses Arabic manuscripts, early printed books, and archival documents in Ottoman Turkish, Arabic, and Persian.

We are proud to announce our strategic collaboration with INSAIT - Institute for Computer Science, Artificial Intelligence and Technology that will enhance the open access to the valuable treasures and knowledge stored at our library. In this line of thinking let me introduce Mr Emilyan Pavlov from INSAIT, who will present more about our partnership.

My name is Emiliyan Pavlov, I am a data scientist at INSAIT and I will be presenting on behalf of the National Library "Cyril and Methodius" in Sofia, Bulgaria. I will tell you about our collaborative projects. They include Automated Metadata Generation, Enhanced Semantic Catalogue Search, as well as digitising and translating Ottoman Archives. About the Institute: INSAIT specialises in Research in Computer Science and Artificial Intelligence. Scientific works every year are accepted in top conferences such as NeurIPS, ICML, ICLR, CVPR. Research fields more in particular are in Trustworthy AI, Quantum Computing, Computer vision which includes robotics and autonomous driving as well as generative AI which is most relevant here.

In 2023 we released the first Open AI Model for Bulgarian Language which specialises in both Bulgarian and English. As you may know, teaching a foundational large language model, makes it forget things, and our method specialises in teaching it new things while mitigating this behaviour. It is available on bggpt.ai where it is used by hundreds of thousands of users. The models behind the project are all available on HuggingFace and I have linked the scientific work behind it. It is already being integrated into many institutional projects in Bulgaria involving the national revenue agency, as well as some municipalities in the form of AI assistants with data that they find relevant.

Cataloguing is a very language specific task so BgGPT is useful for enabling a lot of such downstream tasks. Imagine some entries are missing core fields such as genre and we want to enhance the entry by using augmented generation with web search for generating things like summaries, keywords and target audience. We are proposing to incorporate this with a human-in-the-loop in the following way: If any core fields are missing, we generate them along with the additional ones and give it to a cataloguer for review. If only the auxiliary fields are missing, the entry is forwarded to a specialised database for downstream tasks. An example would be enhanced semantic search.

The already enriched data fields are represented semantically through embedding vectors in the specialised database. That enables us to shift from the regular Lexical Search using algorithms such as BM25 to a Hybrid Search which also takes into account the thematic context of the user's inquiry.

For example, the user is looking for adventure books for teenagers about overcoming injustice. Regular exact matching might match on adventure in the genre field, teenagers in Target audience or injustice in Keywords. However, the semantic representation of context rich fields would rank entries more appropriately.

Yet, another downstream task would be using the retrieved results and their metadata in order to make the catalogue speak with an LLM chatbot. Given a query we give it the necessary context and use the LLM to generate answers in natural language, create reading lists, explain connections between the works etc.

Here, for example, we are asking for books about overcoming injustice suitable for high school students. The retrieval part gives the top results to the LLM and the LLM uses that information to give a natural language answer, not specifying only the plot of these books, could also say something about the edition as well as availability.

To summarise the benefits of semantics, search and the downstream tasks it enables:

Users can search with more natural language phrases and find relevant results. Keywords enable discovery with books with similar themes

- One can use it for Educational filtering, based on target audience.
- Multi-Faceted search combining multiple things to look for such as author, period.
- And by selecting a book, most similar ones based on its entries can be listed.

A more challenging project which is still in inception is digitalisation of the world's 3rd largest Ottoman Archive which is held by National Library "Cyril and Methodius" in Sofia. It includes over 1 million documents in Ottoman Turkish, Arabic, Persian – 4000 manuscripts and 2000 historical books.

There are two planned phases. First, we do Document Recognition Research:

- Building upon existing OCR models to handle Ottoman scripts, Arabic and Persian
- Research approaches for manuscripts variations and historical handwriting styles
- Explore methods to process non-trivial elements on the manuscripts such as decorations and marginalia
- Step on state-of-the-art models for degraded document processing

Then, once we have processed the texts successfully, we can explore methods for translation.

We should research Ottoman Turkish to Bulgarian or English translation approaches, explore domain adaptation for the different types of texts as they vary by legal, administrative and historical texts. Then we can build on multilingual models to enable cross-referencing

Carlo Blum:
Automating Cataloguing in the Age of Digital Legal Deposit: An AI-Based Approach at the National Library of Luxembourg

Abstract

The exponential growth of digital legal deposit collections has significantly impacted national libraries, creating cataloguing backlogs that challenge traditional processing workflows. This paper examines the potential of artificial intelligence (AI) for automated cataloguing, focusing on a case study presented at the 2025 Conference of European National Librarians[1] (CENL). The study outlines a methodology for evaluating AI-generated bibliographic records using field-level and full-record assessments, identifies gaps such as missing subject headings and classification data, and proposes a collaborative, transparent approach to AI training and deployment. The analysis highlights both opportunities and challenges of integrating AI into library cataloguing workflows, emphasising the importance of community-driven solutions to maintain control and align with professional metadata standards. Recommendations for future development and policy considerations are presented.

Keywords:
National Library of Luxembourg, Digital legal deposit, automated cataloguing, MARC 21[2], RAMEAU[3], Dewey Decimal Classification[4], artificial intelligence

Introduction

The digital transformation of publishing has dramatically changed the way knowledge is produced, disseminated, and preserved. National libraries, which have long played a central role in safeguarding cultural and intellectual heritage, now face the task of processing vast amounts of digital material. Legal deposit, the requirement for publishers to submit copies of their works, has historically focused on print publications. However, with the shift to electronic formats, digital legal deposit has become a dominant component of library acquisitions.

This evolution brings both opportunities and challenges. While digital publications offer improved accessibility and preservation potential, their sheer volume and diversity complicate cataloguing workflows. Traditional manual cataloguing, while highly accurate, cannot scale to meet this demand. This has resulted in substantial processing backlogs at the National Library of Luxembourg (BnL) —

[1] https://www.cenl.org/, Last visited: 5 December 2025

[2] https://www.loc.gov/marc/bibliographic/, Last visited: 5 December 2025

[3] https://rameau.bnf.fr/, Last visited: 5 December 2025

[4] https://en.wikipedia.org/wiki/Dewey_Decimal_Classification, Last visited: 5 December 2025; M.P. Satija, The Theory and practice of the dewey decimal classification system. Second edition. Philadelphia, PA, Chandos Pub, 2013.

74,756 unprocessed digital files in the case examined here —highlighting the urgent need for innovative, automated solutions.

Artificial intelligence (AI) offers a promising path forward. By leveraging machine learning and natural language processing, large language models can generate bibliographic records automatically, accelerating processing and reducing reliance on human intervention.[5] This paper explores an AI-based approach to automated cataloguing, known as AutoCat at the National Library of Luxembourg, and outlines its methodology, evaluation process, key challenges, and roadmap for future development.

This development should also be viewed in the context of existing projects like Suominen et al. (2022) on Annif[6] and Finto AI, Ex Libris Group (2024) on the Alma AI Metadata Assistant[7], Bolšteins (2025) on automated retrospective cataloguing at the National Library of Latvia[8], Lowagie (2024) on AI-powered cataloguing at the National Library of Belgium[9] and Labo de l'Abes (2025, avril) *Étude expérimentale de l'indexation RAMEAU assistée par IA*. ABES[10] — which collectively examine recent advances and practical implementations of artificial intelligence for automated bibliographic description and subject indexing in national and academic library settings.

Digital Legal Deposit: Background and Challenges

What is Digital Legal Deposit?

Legal deposit is a statutory obligation requiring publishers to submit copies of their publications to designated libraries. Traditionally, this obligation applied to physical media such as printed books and newspapers. With the advent of digital publishing, legal deposit has been extended to include electronic resources, such as e-books, online journals, and websites.

In Luxembourg, Article 1.3 of the Règlement grand-ducal du 6 novembre 2009 relatif au dépôt légal[11] explicitly states:

---

[6] Suominen, Osma, Juho Inkinen, and Mona Lehtinen. "Annif and Finto AI: developing and implementing automated subject indexing." Bibliographic control in the digital ecosystem.-(Biblioteche & bibliotecari, 2612-7709; 5) (2022): 265-282

[7] Exlibrisgroup. AI Metadata Assistant Preview Initiated for All Alma Customers, 2024, https://exlibrisgroup.com/announcement/ai-metadata-assistant-preview-initiated-for-all-alma-customers/, Last visited: 5 December 2025

[8] Bolšteins, Matīss. National Library of Latvia. Automated Retrospective Cataloguing at the National Library of Latvia (NLL), https://www.cenl.org/wp-content/uploads/2025/02/Automated-ReCataloguing-at-NLL.pdf, Last visited: 5 December 2025

[9] Lowagie, Hannes. National Library of Belgium. AI-Powered Cataloguing, Facet Publishing, ISBN 9781783308071

[10] https://abes.fr/wp-content/uploads/2025/04/etude-experimentation-indexation-rameau-assistee-par-ia.pdf, Last visited: 5 December 2025

[11] https://legilux.public.lu/eli/etat/leg/rgd/2009/11/06/n8/jo, Last visited: 5 December 2025

"The following publications published on national territory are subject to legal deposit in favour of the National Library:

[…]

Publications without physical media made available to the public via an electronic network, in particular websites and Internet content, as well as all signs, signals, writings, images, sounds or messages of any kind, including all publications referred to in the two preceding paragraphs."[12]

Impact on the BnL

Digital legal deposit collections grow rapidly, driven by low publication costs and the ease of digital distribution. For the BnL, this creates four key challenges:

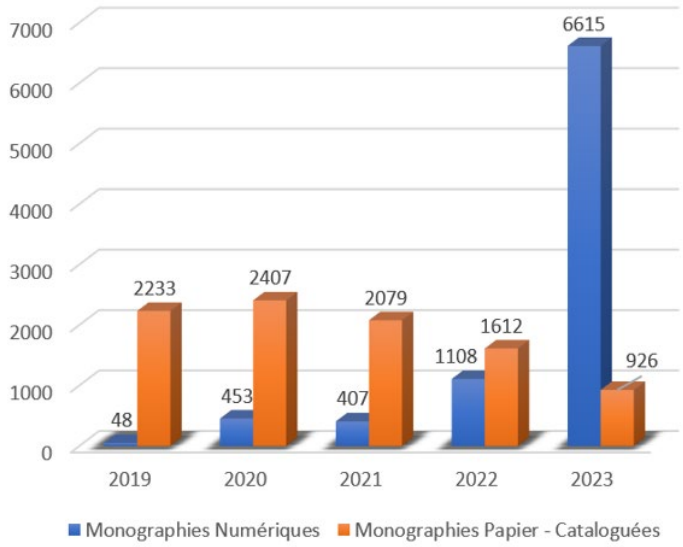
1. **Volume:** Massive influx of files exceeds the capacity of traditional cataloguing teams.
2. **Diversity:** Files come in different formats (PDF, EPUB, HTML), requiring specialised handling.
3. **Metadata Complexity:** Digital items often lack standardised metadata, requiring libraries to extract and normalise information.
4. **Layout Variability:** Standard layouts of print publications do not carry over into the digital realm and concepts such as cover or title pages, publication details etc. are often not included by authors and publishers.

These cumulative effects created cataloguing bottlenecks, with tens of thousands of items awaiting processing. This situation limits timely access and discoverability, effectively preventing patrons from using a substantial part of the library's collection. As a result, the library is unable to fully deliver on its mission to provide comprehensive and equitable access to knowledge, which ultimately represents a poor level of service to its patrons.

[12] "Sont soumises au dépôt légal en faveur de la Bibliothèque nationale les publications suivantes éditées sur le territoire national:

[…]

Les publications sans support matériel mises à disposition du public à travers un réseau électronique, notamment les sites et contenus Internet, ainsi que tous les signes, signaux, écrits, images, sons ou messages de toute nature, y compris toutes les publications visées aux deux paragraphes précédents."



Metadata standards and their Importance

Automated cataloguing capture descriptive metadata (e.g. title, author) but could also assign subject headings and classification codes accurately,[13] a task traditionally reliant on expert cataloguers. Since the BnL uses MARC21 as its standard for bibliographic records, AutoCat must be capable of producing output in this format.

MARC 21: The foundation

Like many libraries, the National Library of Luxembourg uses MARC 21 (MAchine-Readable Cataloging) to encode metadata elements such as title, author, publisher, publication date, subject headings, and classification numbers.

However, MARC 21 by itself does not define how cataloguers should describe resources — it only defines how to encode data once it is created. This descriptive guidance comes from RDA[14] (Resource Description and Access). RDA provides instructions for describing and identifying resources and their relationships, supporting linked data and user-focused discovery.

Together, RDA and MARC 21 form the foundation of the BnL's bibliographic control:

- **RDA** provides the content rules — what information to include and how to formulate it.
- **MARC 21** provides the encoding format for representing that information in library systems.

This combination ensures both human-readable clarity and machine-actionable metadata, enabling interoperability across library catalogues and supporting long-term digital preservation and access.

[13] Cf. Christoph Poley Sandro Uhlmann, Frank Busse, Jan-Helge Jacobs, Maximilian Kähler, Matthias Nagelschmidt, Markus Schumacher: Automatic Subject Cataloguing at the German National Library. In: LIBER quarterly 35-1 (2025), 04, p.1-29

[14] https://www.rdatoolkit.org/

Subject Indexing: RAMEAU and Dewey

Automated cataloguing must not only capture descriptive metadata but also helpful to assign subject headings and classification codes accurately, a task traditionally reliant on expert cataloguers. At the BnL, this process involves multiple classification systems:

- **RAMEAU** (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) is a French subject authority file widely used for subject indexing. It encompasses approximately 209,000 authority records in the Luxemburgish union catalog.
- **DDC** (Dewey Decimal Classification) provides a hierarchical numeric system for organising knowledge, enabling patrons to locate materials by subject.
- **Internal subject classification** (a library-specific taxonomy composed of 70 categories)
- **National DDC based bibliographic code**

AI for Automated Cataloguing (Autocat)

BnL AI Project Development Cycle

Over the past five years, the BnL has conducted a series of artificial intelligence (AI) projects within an experimental framework. These projects are conceived as exploratory pilots aimed to evaluate the potential of emerging technologies for library operations. The BnL also acknowledges that these projects may be discontinued at an early stage if meaningful outcomes are not achieved. Each initiative typically spans a six-month project cycle and is implemented by a compact, interdisciplinary team composed of a half-time AI specialist, an information technology expert dedicating 25% of his time, and one subject domain expert. Some of the results of these projects are available on the BnL's Github[15].

The AI-assisted cataloguing initiative follows this cycle, ensuring that the resulting tools are both functional and aligned with library needs. The process includes the following stages:

Identifying the Goal
The first step is to clearly define the objectives of the AI project, such as reducing cataloguing backlogs, improving metadata quality and accelerating workflow efficiency.

Selection of an AI Method
Based on the defined goals, appropriate AI methods are selected.

Prototype Development
A prototype is built using a representative dataset, allowing initial testing of the AI model's capabilities.

[15] https://github.com/natliblux/

Gathering Feedback from Key Users

The prototype is shared with librarians and its-experts, who evaluate the results produced to its usability, performance, and compliance with metadata standards.

Adjusting the Approach

Feedback is shared and, if necessary, used to refine the AI model and workflows. This iterative process continues until satisfactory performance levels are achieved or the project is discontinued as no promising result is produced.

Integration into Existing Systems

Once validated, the AI solution is integrated into the library's digital infrastructure, ensuring interoperability with cataloguing system (Alma[16]) of the National Library of Luxembourg, its digital preservation system (Preservica[17]) and its presentation platform for digital content (eluxemburgensia[18]).This work is done by the BnL's IT team responsible for development on the production systems.

Communication and Promotion

The results and benefits of the AI project are shared with stakeholders, both internal and external, to encourage adoption and transparency.

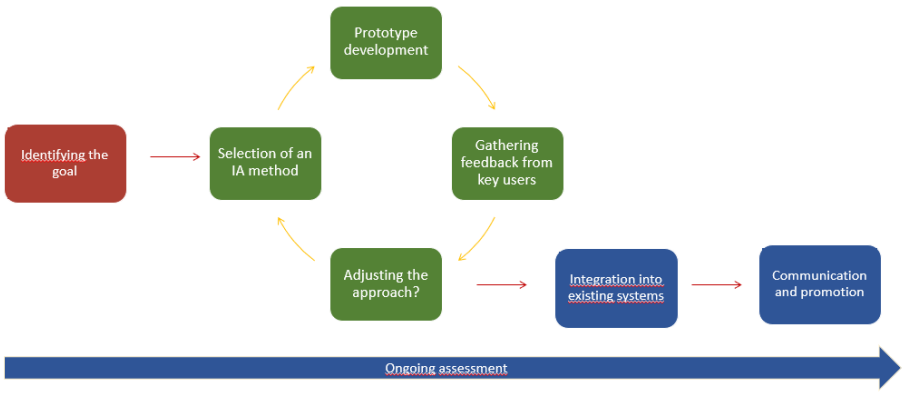


Image 1: AI Project Development Cycle

[16] https://www.a-z.lu, Last visited: 5 December 2025; https://exlibrisgroup.com/products/alma-library-services-platform, Last visited: 5 December 2025

[17] https://preservica.com/, Last visited: 5 December 2025

[18] https://eluxemburgensia.lu, Last visited: 5 December 2025

Overview of the Approach

Introduction

The AutoCat project uses AI models to automatically generate metadata records, focusing in a first phase on descriptive fields. In a second phase subjects and classification should be included. The first phase workflow involves:

1. Extracting text and metadata from digital files.
2. Using AI to predict bibliographic elements.
3. Exporting results in MARC 21/RDA format for library system ingestion (ALMA).

Detailed overview

The workflow begins with the ingestion of a digital document, which includes both image-based content and extracted textual data. This initial step ensures that all relevant visual and textual components are available for subsequent processing and analysis.

A finetuned GPT-4o Vision model from OpenAI[19] is used to analyze the images contained in the document to derive the following MARC21 fields:

- **100, 110, 245, 264, 300, 700, 710** (author, title, publication, and physical description).

The GPT-4o Vision model was finetuned using 300 ground truth pairs, each consisting of a full text document and it's corresponding manually curated and validated bibliographic record.

The textual content of the document is processed through the following steps:

- Content summarisation: A concise summary of the document is generated using GPT-4o Mini.
- Language detection: All languages present within the text are identified using Lang-ID, a tool developed by the BnL.
- ISBN extraction: ISBN numbers are detected through the application of regular expressions.
- National bibliographic code generation: The national bibliographic code, recorded in MARC field 924, is produced using Annif[20].

All extracted metadata undergoes post-processing for normalisation and alignment to RDA-compliant cataloguing rules.

The processed metadata is explicitly marked "AI generated" and exported as MARC21 XML, ready for integration into the library catalogue (Alma).

[19] https://openai.com/

[20] Annif. *Annif – Automated Subject Indexing for Libraries*. Accessed July 25, 2025. https://annif.org

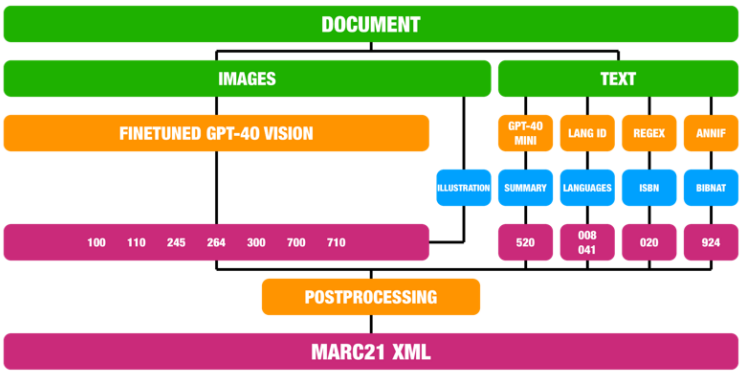


Image 2: IA workflow for cataloguing: document images and text are processed into MARC 21 XML using GPT-4o models and library tools.

Evaluation Methodology

To check the quality of the AutoCat generated bibliographical records, the results produced must be evaluated. The objectives of this evaluation are:

- to measure the overall quality of the generated records,
- to determine which types of errors can be considered acceptable and which are unacceptable, and
- to evaluate whether AutoCat can operate fully automatically, without human intervention.

A major challenge of this project is the implementation of a structured evaluation framework, designed to assess AI performance systematically. The BnL decided to evaluate the records using a two-level approach: assessing the accuracy of each individual field, as well as the quality of the record as a whole.

Field-Level Evaluation

Each bibliographic field — title, author, publication date, etc. — is compared to the ground truth by a domain expert. This allows granular assessment of which elements AI handles well and where it fails. For example:

- **Title accuracy:** Did AI correctly identify and transcribe the title?
- **Author identification:** Did AI correctly distinguish primary and secondary authors?

Full-Record Evaluation

Beyond individual fields, the coherence of the record as a whole is examined. Does it provide a usable description of the resource?

Scoring System

A qualitative scoring system classifies results into four categories:

- **Perfect:** Identical to ground truth.
- **Okay:** Minor errors that do not impede use.
- **Bad:** Significant inaccuracies requiring human correction.
- **Unusable:** Output unsuitable for cataloguing.

This dual perspective ensures both precision and practical usability are considered.

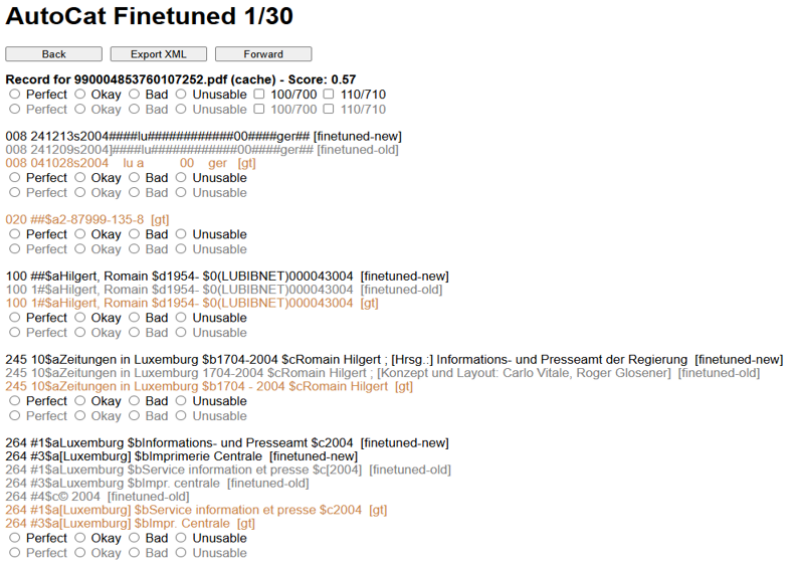
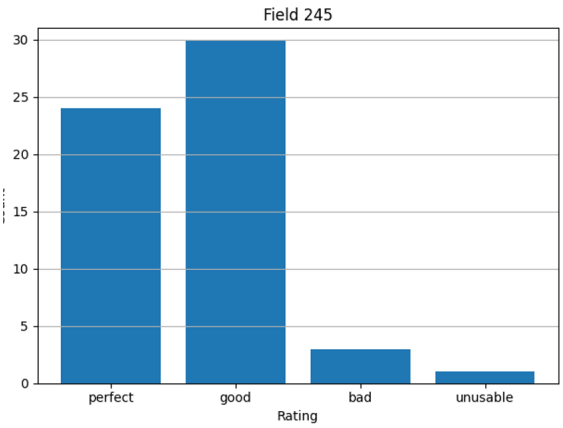
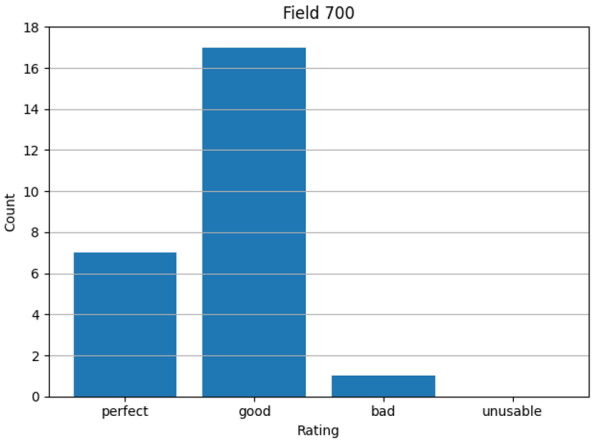


Image 3: Screenshot record evaluation tool

The following two diagrams illustrate selected findings from the BnL's analysis.



Evaluation of Field 245 (Title Statement) produced by AutoCat, showing the distribution of quality ratings. Most records were rated as "good" or "perfect," while only a few were considered "bad" or "unusable."



Evaluation of Field 700 (Personal Name) produced by AutoCat. Most records were rated as "good," with several rated as "perfect," while only one was considered "bad" and none were "unusable."

Challenges Identified

The evaluation shows promising results but needs to be conducted on a larger dataset. The National Library of Luxembourg aims to evaluate 1,000 records.

It is evident that there are significant differences between AI-generated (AutoCat) records and fully validated bibliographic records.

- Missing subject headings:
  AutoCat did not yet assign RAMEAU terms. This is significant because subject indexing supports thematic discovery and intellectual access.
- Missing classification:
  Dewey Decimal Classification (DDC) integration is pending.
- Poor generation of the national bibliographic code (920):
  Essential for national bibliographies, this element generated but needs improvement.

Automatically generating RAMEAU subject headings and Dewey Decimal Classifications is challenging because training an effective ANNIF model requires a large set of ground truth data, which the National Library of Luxembourg does not currently possess. Both indexing and classification systems encompass numerous concepts and demand extensive training data to achieve accurate results. Furthermore, given that the collection of the National Library of Luxembourg is multilingual, language introduces an additional layer of complexity.

This leads to the following conclusion: national libraries possess extensive sets of validated metadata and full-text resources that can serve as valuable training data for AI models, such as those designed to automate indexing processes for example. Individual libraries have the capacity to train these models locally and subsequently share the trained models within the library community. This

approach facilitates the distribution and reuse of AI models across institutions, fostering collaboration and efficiency.

Such a collaborative framework offers multiple advantages:

- Shared development costs: By distributing the workload and expenses, the financial and technical burden on individual institutions is reduced.
- Diverse training data: Pooling multilingual and specialised collections enhances the robustness and adaptability of AI models, improving their capacity to process heterogeneous data accurately.
- Transparency and trust: Libraries maintain control and oversight over both the training data and algorithmic processes, ensuring ethical standards and institutional values are upheld.

Roadmap and Goals

To address these gaps, the project has established clear objectives:

1. **Evaluate 1,000 documents:** This sample size should be sufficient to understand model performance.
2. **Optimise evaluation speed:** Automation of scoring will allow faster model iteration.
3. **Expand training data:** A larger, more diverse dataset will improve AI performance across genres and formats.
4. **Version 1 Release:** The goal is to deliver a functional AutoCat system by the end of 2025.

These milestones balance immediate cataloguing needs with long-term AI capability building.

Conclusion

The growth of digital legal deposit has outpaced traditional cataloguing methods, creating backlogs that threaten access to information. AI-assisted cataloguing offers a promising solution, capable of processing large volumes of digital files rapidly and accurately for descriptive metadata. However, challenges remain in subject indexing, classification, and compliance with the BnL's standards.

The evaluation methodology adopted here — combining field-level and full-record assessments with a qualitative scoring system — provides a replicable framework for improving AI performance. Achieving a fully functional AI-driven workflow by the end of 2025 will depend on continued development, expanded training data, and collaboration across libraries.

The BnL is confident that AI will not replace human cataloguers but rather transform their roles:

- **From routine data entry to quality control and exception handling.** Human cataloguers will increasingly focus on supervising AI outputs, verifying accuracy, and handling exceptions where judgment, domain knowledge, or contextual understanding is

required. This shift allows cataloguers to spend less time on mechanical entry and more on ensuring data quality, consistency, and adherence to standards.

- **From descriptive work to curatorial and subject-specific expertise.** Traditionally, cataloguers spend significant effort describing the content of items, often following strict metadata schemas. With AI assisting in generating initial descriptions and subject headings, human experts can dedicate more attention to interpretive and curatorial work. This includes applying specialised knowledge to enrich metadata in ways that improve discovery and access. Essentially, AI can help to make library collections more meaningful, discoverable, and aligned with patron needs.

Commercial AI solutions often rely on proprietary training data, limiting libraries' insight into how records are generated. This raises concerns:

- Data governance: Libraries may not control how their data is used.
- Bias and accuracy: Proprietary models may not align with local cataloguing rules.
- Long-term sustainability: Vendor lock-in may reduce flexibility.

Hence, automating cataloguing is a compelling case for libraries to **collaborate and to share trained models**. This collaborative approach also mirrors historical library initiatives, such as shared catalogues and cooperative metadata standards, and extends them into the AI era.

By embracing innovation, libraries can use AI technologies to manage digital legal deposit effectively, ensuring timely access to knowledge for their patrons.

Renata Sadunišvili:
Book Recommendation Tool for Public Library Readers in Lithuania

Why we do it: Increasing challenges of the competitive environment for libraries and the expectations of users affected by them

When thinking about the reader's path to a book in a public library, we often see the following scenario: the user already knows what work or author he wants to read, finds the publication in the Lithuanian Integrated Library Information System (LIBIS) and orders it. Sometimes they consult with librarians, who, based on their experience or recommendations from other readers, advise what to choose. It is easy for a reader who knows what he is looking for to quickly find a publication in the digital catalogue. But what to do when you don't know what to read?

Then the book reader's "field of view" expands and other book agents appear in it, who communicate significantly more clearly and actively about books. They have a significant influence on reading choices. Publishers and bookstores (physical or virtual) become such agents. The lists of the most popular books that they compile, extensive visualisations, detailed descriptions of the book content, recommendations from various professionals or simply famous people, and sales charts serve very well to popularise books as a category. Although the range of publications they have is much narrower than that of libraries, considering that a non-professional reader reads up to 20 books on average per year, this offer in any case seems to exceed the real options for choice.

Many promotional measures used by bookstores and publishers are focused on presenting new works. This is consistent with their business model, but looking more broadly, by constantly emphasising the search for innovations in the book selection process, the user, who passively observes the literary field, such an emphasis forms an expectation that is not always favourable to libraries: to choose a new publication, even if that "new work" is just another edition of The Little Prince.

This approach poses challenges for libraries, as they are not simply "agents of new works" and their funds are incomparably more comprehensive, covering a wider range of topics and a longer publication period. The commercial reason behind is also no less important. Libraries, financed by the public, must balance the budget of their procurement and see a longer perspective than 1.5 years. This is how long, on average, the active demand for a new publication remains, after which it irreversibly subsides, leaving the repositories filled with unreadable copies.

Currently, the search tool available to the reader of Lithuanian libraries is not very competitive. Due to copyright restrictions, bibliographic records in the digital catalogue are rarely supplemented with cover photos, lack a description of the publication's content, and a significant part of them is more oriented towards the academic user. In other words, libraries are losing this battle of "brightening of

new works", failing to respond to the changed expectations of users to choose from more visualised content supplemented with summaries and recommendations.

Despite objective thresholds, libraries have no less data resources, on the basis of which they can provide recommendations no worse than the private book sector.

The information accumulated in the shared LIBIS catalogue reveals what readers of all Lithuanian public libraries read, regardless of which library they borrowed the book from. By analysing this data, it is possible to compile lists of the most popular authors, ratings of the most read books, compare reading preferences of regions or age groups, and monitor their changes over time. By aggregating LIBIS data, it becomes possible to form recommendations for those readers for whom reading is a form of leisure, not part of professional activity.

The National Library, in order to improve the organisation of reading services and having assessed the limitations of the existing search tool offered to library users, is developing a book recommendation system based on internal big data. This system will help to redistribute the flow of users, expanding the range of offers and shifting the emphasis away from publishing news.

For users who do not have lists of literature or authors that meet their interests or professional needs, and who read non-systematically, we plan to offer a tool that will make book searches more efficient, based on the experiences of other readers. This tool will also be useful for librarians who are asked the question - what would you recommend? Unlike publishers and bookstores, the rating provided by the library will not be focused solely on newly published books. It will include everything that has recently been interesting to readers in Lithuania. And this is almost one sixth of the population of Lithuania, each of whom reads about 15 books per year, borrowed from 1208 public libraries.

Who we collaborate with: A partnership between the National Library and the Lithuanian State Data Agency in securely working with data and effectively transforming it into consumer-centric products

Currently, the Department of Digital Transformation Development of the National Library, in collaboration with the Lithuanian State Data Agency (SDA), is developing a book recommendation tool.

The data of the shared LIBIS catalogue, which includes book borrowing records, is transmitted to the State Data Agency every month. This obligation applies to all Lithuanian state institutions – from healthcare to transport. Employees of institutions can work with anonymised copies of data on the Palantir platform administered by the SDA. It allows you to work effectively not only with your own data, but also enrich it with copies of anonymised data from other institutions. The SDA is responsible for the anonymization of all data and compliance with other security protocols so as not to violate the GDPR and other requirements set out in the legislation.

Transforming data from tables and static diagrams into more convenient formats – scoreboards or apps that are more suitable for the daily use of a library customer – requires not only specific knowledge and skills, but also constantly changing tools. Large-scale data processing, merging, analysis and visualisation of different formats have long become separate professional areas. Cultural and memorial institutions often lack specialists of various profiles who can work effectively and productively with their data. Since such tools evolve quite quickly and their licenses are expensive, the opportunity provided by the state to all institutions of public sector to use a unified platform enables the use of institutional data in the creation of services and products for public needs. This provides state institutions with the opportunity to conduct broader insights, plan activities more precisely, form strategies, and create more complex products. In this context, libraries become full-fledged participants in the national data ecosystem.

By cooperating with the State Data Agency, the National Library receives not only infrastructure for data storage and analysis, but also competencies that are difficult for libraries to achieve independently: secure data anonymization solutions, access to a specialised environment for work with data, the ability to create visualisations and complex analytical models.

The National Library brings unique content expertise to the cooperation – deep knowledge of bibliographic records, publishing history, and reading processes. It is the library specialists who help to define precisely which links between the data are meaningful and which are just statistical noise. Therefore, the project is the result of mutual synergy: SDA provides technological tools, the National Library – cultural and professional content.

How we do it: The process and challenges of developing a work recommendation tool

The output of the book recommendation tool for the user is simple - it resembles the usual lists of offers generated in e-stores based on order history. In other words, the list of recommended books is compiled from the works that library visitors have read, as well as replacement for the work that we just read (Fig. 1). If the set of readers A and the set of readers B both read the same book, it is likely that the set A may also be interested in other books read by the set B.

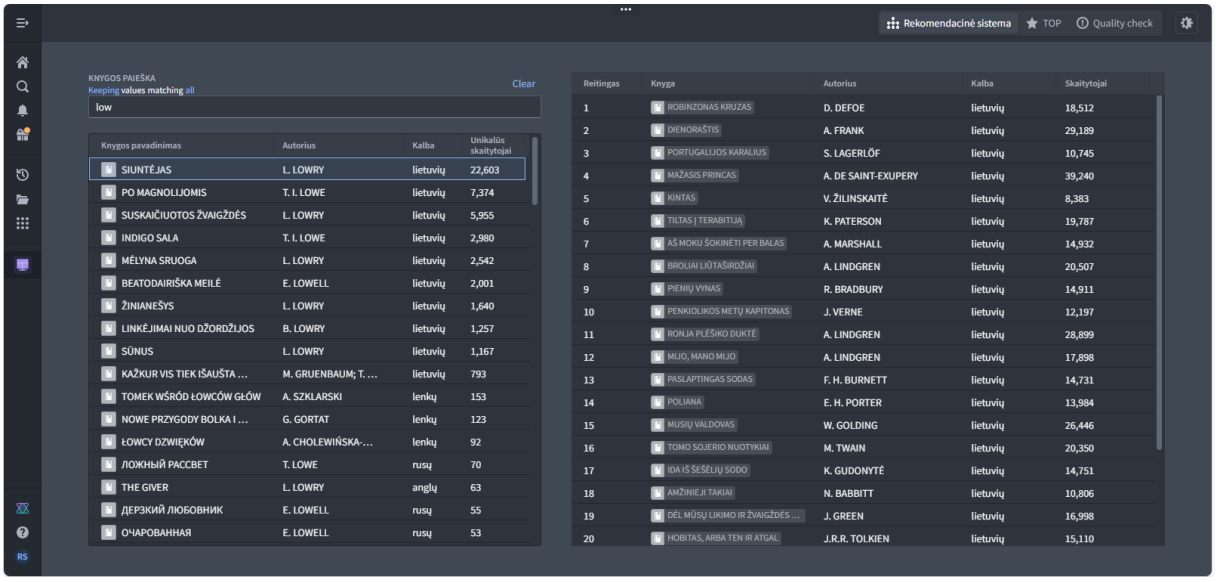


Figure 1. Prototype interface of the work recommendation system

After starting the data preparation work, a well-known truth was confirmed: even when the data is collected very neatly and correctly (for example, bibliographic records), their formats are often not suitable for analysis on platforms designed for working with big data. In addition, the data analysis tools available to organisations themselves are usually not adapted to work with large data sets located on different platforms. In this case, cooperation with the SDA became a reasonable decision that allowed us to keep up with the changes of time and technology.

When formatting the data, it turned out that the Universal Decimal Classification (UDC) can become a useful additional layer in forming the sample and separating groups of records within the sample. From the Shared Catalogue LIBIS, which is used by all public libraries in the country, a separate sample was isolated according to the format of the publication, in this particular case, a printed book. This sample was divided into two subgroups using the UDC data of the publication - fiction and non-fiction.

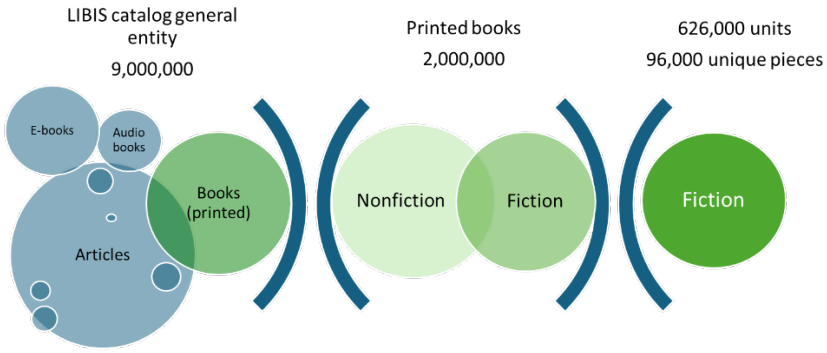


Figure 2. Documents being recorded in LIBIS bibliographic records. LIBIS data

Currently, work is being carried out on the entire collection of printed fiction books. After testing the model, its algorithm will be applied to the larger collection of printed non-fiction works.

In order to form recommendations not at the level of a single printed copy, but at the level of the relevant work, it is important to correctly link the borrowing tracks of all editions of that work (which can be up to 100) and all copies of editions (which can be significantly over 100) over a certain period of time. Therefore, first of all, the catalogue entries were "compressed": copies were reduced to the publication, and publications to the work. At this stage, it was necessary to take into account not only the number of different editions of the same work, but also its editions in foreign languages. Lithuanian library users often read in several languages, and their "basket" includes publications in English, Russian, Polish, German, French, etc., and readers can borrow the same work in different languages. Taking this into account, algorithms were developed that allow combining works published in different languages into a common network of recommendations. Without taking linguistic diversity into account and without assessing the importance of this factor in the selection process, some of the links between books would be "erased" and recommendations would become monolingual.

When aggregating bibliographic records, it was important not to lose the borrowing "tracks" associated with copies and their links with other books borrowed by the same user. Borrowing data in the system is linked to a specific copy, but when recommending a work, it is important to take into account all track records of a reader so that the offer would include both the borrowing links of all copies with other copies of the same work, as well as copies of other works borrowed by the same user, which were borrowed by other users. The list of offers is formed based on the borrowing history of all works available in libraries, using only transaction records, since LIBIS does not record the search path.

The second important stage was the development of a mechanism for matching of last names of authors. Depending on the historical period of the Lithuanian state in which the book was published, the name and last name of the same author may be written differently in the bibliographic records of the LIBIS catalogue: with or without Lithuanian endings, in the original language, with or without initials. Since the catalogue records the actual information provided on the book cover, in order for the system to understand that all these entries refer to the same author, it was necessary to create a mechanism for matching author names, so that when foreign author names are entered in a language other than the original, the system would recognise them and find the work that the reader has in mind. This problem is less relevant for Lithuanian last names, but it is extremely important, for example, in the case of French or English last names, which were written differently on book covers by Lithuanian publishers in different historical periods, this is important. The recommendation lists will be formed based on all transcriptions of the author's last names of the work, regardless of the year in which, for example, "The Little Prince", read by the reader was published.

Another equally important stage was the recognition of books included in general education school curricula and their weight reduction in the recommendation lists. These works are more often a required reading at a certain stage of learning, rather than an individual choice of the reader. Their borrowing statistics are significantly higher than average, so leaving them in the recommendations as equivalent to other works would disproportionately skew the proposals towards general education programs, which would likely not meet the needs of the non-school audience. If an official list of such

works were available, it would be easy to adjust the weight of these works, but currently there are only recommendations that mostly indicate authors, not specific works. Trends in borrowing data over a longer period helped to distinguish school works. After the observation of the borrowing dynamics over time, it can be seen that the borrowing trend of books that fall into general education programs is seasonal, with its peaks each year. Meanwhile, other books reach their borrowing peak within the first two years of publication and then usually do not repeat it.

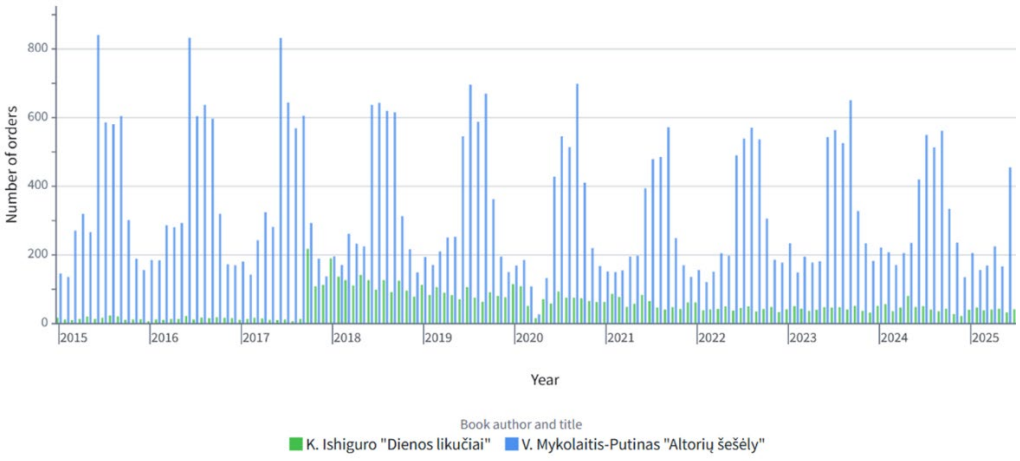


Figure 3. Trajectory of borrowing of school and non-school works over time. LIBIS data

Results and perspectives

A prototype of a recommendation tool for fiction is working. By entering a book title or author's last name in the search box, entries related to the entered keywords are displayed. By selecting one of them, a list of other books is generated - those that users have also read, after reading the selected work. In addition, it is possible to see monthly updated general TOP 25 lists compiled according to the choices of readers of all Lithuanian public libraries, as well as by individual age groups.

This project is not only a technological experiment, but also a cultural mission. Its goal is to help Lithuanian readers discover more authors and their works, and for libraries to become even more important participants in modern reading culture.

Tamara Butigan Vučaj:
LibrAIfy: Unlocking Serbian Written Heritage with AI

LibrAIfy is a result of the very first AI project undertaken by the National Library of Serbia, developed by the Mathematical Institute of the Serbian Academy of Sciences and Arts and supported by the Serbian Government through its Innovation Fund programme *GovTech*.

LibrAIfy is an AI-powered pipeline designed to enhance the digitisation of historical periodicals by unlocking text from digital images. The solution is tailored to the digital library platform of the National Library of Serbia[1] and is built on open access technologies, libraries and AI models in computer vision and natural language processing (NLP). Developed in Python, the system is designed to evolve and improve as more high-quality documents become available.

The Serbian government allocated funding and launched the pilot *GovTech* programme in late 2023 through its Innovation Fund. The programme aims to encourage the public sector in Serbia to adopt innovative solutions, accelerate digital transformation and embrace disruptive technologies, including AI. The National Library of Serbia recognised this opportunity and applied to the programme with a challenge: improving OCR for the Serbian language, particularly for digitised 19th- and 20th-century public domain content. The Mathematical Institute of the Serbian Academy of Sciences and Arts, a respectable scientific institution and also long-term partner of the National Library of Serbia in building its Digital Library, joined the *GovTech* programme to provide a solution to this challenge.

Why Was OCR Improvement Necessary?

Serbia, a South eastern European country with a rich tradition in culture, including the written culture, with the oldest manuscripts dating back to the 12th century. Historically straddling East and West, geographically, politically, and culturally, Serbian language uses both Latin and Cyrillic scripts. This dual-script usage presents a unique challenge for successful OCR, especially in early 20th-century periodicals where both scripts often appear on the same page.

Another major obstacle lies in the quality of the digital images. Many were created in the early 2000s at low resolution, including the newspaper *Politika*, which alone comprises around 700,000 pages. The National Library of Serbia began OCR processing relatively late, starting in 2018 as part of a Europeana project that produced the first searchable PDF documents. However, much of the digital library remains non-searchable. In addition to OCR challenges, complex layouts, historical fonts, and spelling variations in old periodicals further hinder text searchability.

[1] https://digital.nb.rs/, Last visited: 5 December 2025

What Does LibrAIfy Do?

LibrAIfy is tailored OCR solution for Serbian written heritage. It consists of several modules:

- Image enhancement
- Document layout analysis
- OCR processing
- Post-OCR correction
- System performance monitoring

Input documents include:

- PDF files converted to black-and-white JPEG images
- JPEG images from web books available in the digital library

Output formats include:

- .txt
- XML ALTO
- Searchable PDF

LibrAIfy delivers significantly better OCR results compared to existing tools used by the National Library of Serbia, including Tesseract, ABBYY FineReader, and docWizz. The system was trained on four selected periodicals published between 1885 and 1941, using a dataset of 16,000 documents totalling 400 GB.

Module for Automatic Image Enhancement

This module is based on the Restormer[2] (Restoration Transformer) model fine-tuned on the Shabby Pages 2023 dataset. The model performs automatic denoising and deburring of old document images. While the visual improvements may not always be immediately noticeable, they contribute significantly on average, to the overall OCR performance.

[2] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M.H., 2022. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5728-5739)

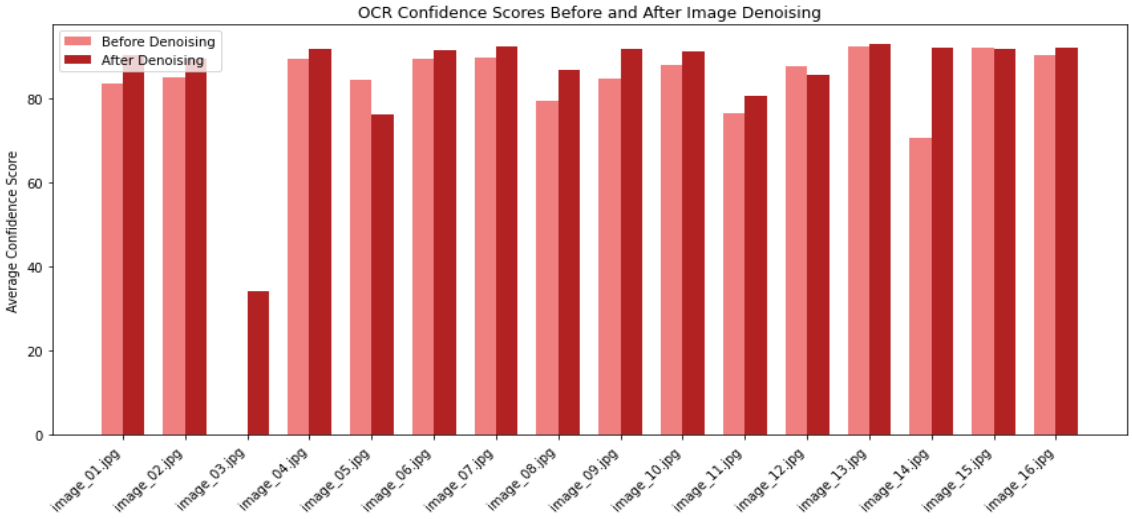


Figure 1. OCR confidence score before and after image denoising

Module for Document Layout Analysis

This module uses the YOLO[3] v8 model, trained on IBM's DocLayNet dataset[4], created in 2022 by the IBM Research team. The dataset contains approximately 80,000 documents with spatial annotations in COCO format, totalling around 28 GB. It includes a wide range of document types such as:

- Financial reports
- Patents
- Scientific papers
- User manuals
- Legal documents
- Tenders

These documents exhibit complex spatial structures. The model supports 11 types of layout elements, including:

- Text blocks
- Titles
- Section headers
- Images
- Tables
- List items
- Page headers and footers
- And others

[3] You Only Look Once: Unified, Real-Time Object Detection, https://pjreddie.com/yolo, Last visited: 5 December 2025

[4] https://huggingface.co/datasets/ds4sd/DocLayNet, Last visited: 5 December 2025

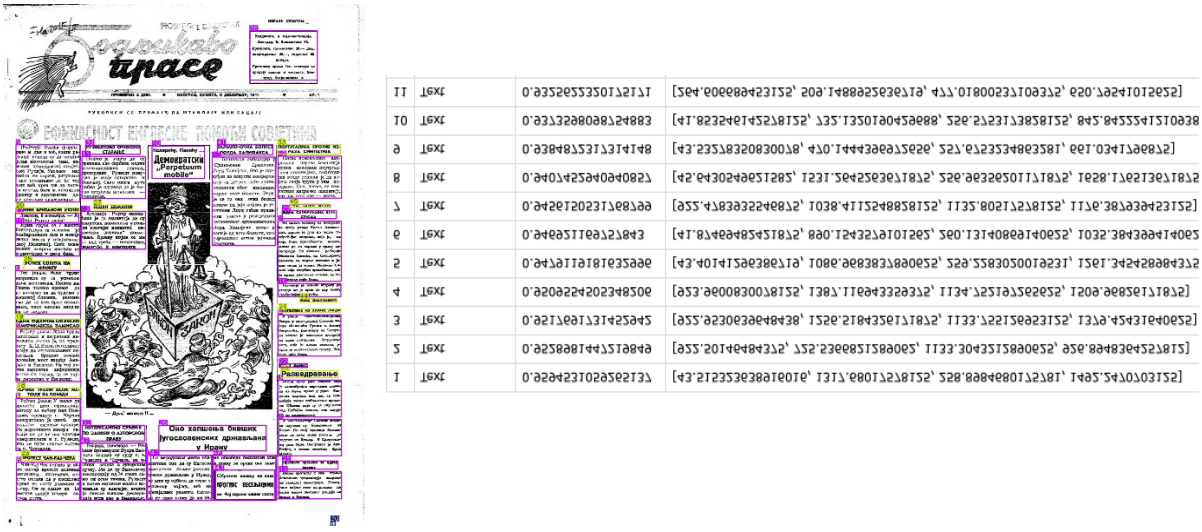


Figure 2. Newspaper page with marked spatial elements and Reliability assessment of the model by element type

Module for OCR Processing

Development began with a review of open-source OCR tools, including:

- OCR4all
- Kraken
- PaddleOCR
- Calamari OCR
- DocTR
- Tesseract

Tesseract v5.5.0 was selected for its robust architecture based on convolutional neural networks (CNNs) and recurrent neural networks (LSTMs), as well as its native support for Serbian Cyrillic and Latin scripts.

OCR accuracy is evaluated by comparing the output text to the correct reference text. This involves tracking the number of edit, insertion, and deletion operations required to transform the OCR output into the expected result. Key metrics include:

- Character Error Rate (CER)
- Word Error Rate (WER)
- Levenshtein Distance

These evaluations require manual verification, which demands significant time and human resources.

Module for Post-OCR Correction

This module is based on a custom neural model for sequence processing. The starting point was the SrpElTec[5] corpus, developed within the COST Action Distant Reading for European Literary History. It consists of digitised and manually corrected 19th-century Serbian novels.

Using this corpus, text transformations were designed to detect and track OCR errors. The final step involves training a recurrent LSTM neural network to map the transformed text back to its original form, essentially performing text denoising.

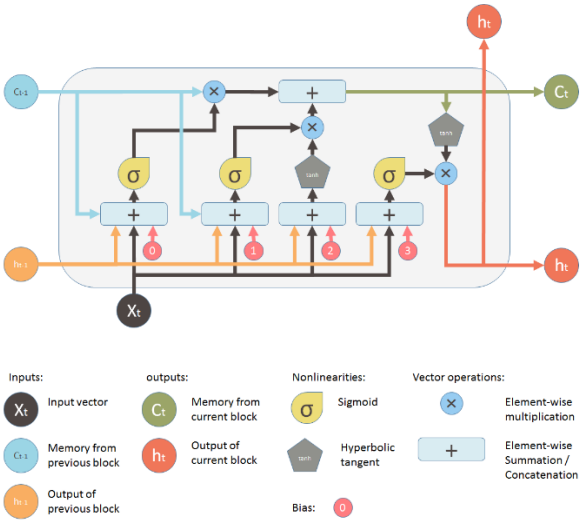


Figure 3. Shallow bidirectional LSTM model with transformed paragraph as input and improved text as output

System Monitoring Module

LibrAIfy also includes a system monitoring module that tracks several types of metrics:

- Dataset statistics
- Module performance statistics
- Computational capacity statistics

LibrAIfy represents a significant step forward in preserving Serbia's written heritage. The project team aims to adapt it for universal use, particularly for small and underrepresented languages.

[5] https://huggingface.co/datasets/jerteh/SrpELTeC, Last visited: 5 December 2025

Ines Vodopivec:
AI4GLAM and CENL: Toward a Global Alliance Transforming Libraries with Artificial Intelligence

Introduction

Artificial Intelligence (AI) is no longer a futuristic concept, rather a present-day reality reshaping every sector, including libraries. As Henrik Skaug Sætra aptly stated in *Technology in Society* (2023), "Artificial Intelligence is here to stay."[1]

Today's new technologies we call AI are making a fundamental shift in how libraries operate and deliver services to their end users. This transformation is already having a significant impact, as institutions actively rethink their workflows to integrate AI-driven processes. This evolution includes identifying gaps and unmet needs within existing systems, equipping staff through targeted education and training to work effectively with AI tools, and recruiting new talent with expertise in data science, machine learning, and digital humanities to build robust internal AI support structures.

These changes are not merely technical — they are strategic. Institutions must adapt their policies and strategies to redesign organisational structures and workflows, reconsider the allocation and use of funds and expenditures, and ensure the ethical implementation of AI. Equally important is the need to foster trust and maintain a positive attitude among employees, while thoughtfully addressing any concerns or fears they may have about the transformation process. In addition, ethical considerations are paramount, especially when AI systems influence public access to information and cultural heritage.

Artificial Intelligence for Libraries, Archives, and Museums' (AI4LAM)[2] core philosophy is simple yet profound: *"Individually, we are slow and isolated. Together, we can go faster and farther."* This collaborative spirit was formalised in 2018 with first Memorandum of Understanding (MoU) signed by six pioneering institutions: The Library of Congress, Smithsonian Institution, National Library of Norway, Bibliothèque nationale de France, British Library, and Stanford University. Since then, AI4LAM has grown into a vibrant community of over 1,500 members by 2025. This expansion reflects a growing recognition that AI can address both internal and external challenges faced by GLAM (galleries, libraries, archives and museums) institutions. From metadata creation to user engagement, AI4LAM's initiatives span a wide spectrum of applications, all rooted in cross-sectoral collaboration. At the 2025 Annual General Meeting of the Conference of European National Librarians (CENL), AI4LAM highlighted the core values of integrating artificial intelligence within the GLAM

[1] Henrik Skaug Sætra, Generative AI: Here to stay, but for good?, Technology in Society, Vol. 75, 2023, ISSN 0160-791X, https://doi.org/10.1016/j.techsoc.2023.1a02372, Last visited: 5 December 2025

[2] AI4LAM, a global community dedicated to integrating AI into the workflows of cultural heritage institutions, ai4lam, Last visited: 5 December 2025

sector and advocated for strengthened international cooperation to transform how institutions manage, preserve, and provide access to cultural knowledge.

Libraries and GLAM in the AI Ecosystem

GLAM institutions play a diverse and dynamic set of roles within the evolving AI ecosystem. Far from being passive recipients of technological change, these institutions are actively shaping the development and application of AI in the cultural heritage sector.

First and foremost, libraries serve as **data producers** and **data providers**, generating and curating vast collections of cultural heritage content. This includes digitised (art)works, historical documents, audio-visual materials, and metadata, all of which form the foundational datasets that fuel AI research and applications.

They also act as **data storages,** responsible for the long-term preservation of legal deposits, digital twins, and harvested web content. Their stewardship ensures that culturally significant data remains accessible, authentic, and usable for future generations and for AI systems that rely on high-quality, well-maintained datasets.

In addition, many GLAM organisations are emerging as **AI developers** themselves. They are increasingly involved in designing and training machine learning models tailored to their specific needs, such as automated metadata generation, image recognition for artworks, or natural language processing for historical texts.

As **AI collaborators**, libraries frequently partner with universities, technology companies, and digital repositories to co-create AI tools and frameworks. These partnerships foster innovation and ensure that AI solutions are informed by domain expertise and ethical considerations unique to the cultural sector.

Finally, libraries are also **end users** of AI technologies. They apply AI tools in real-time to enhance visitor experiences, streamline cataloguing processes, improve accessibility, and uncover new insights from their collections.

This multifaceted involvement underscores the critical role of GLAM institutions in the AI landscape. By contributing data, expertise, and innovation, they are not only adapting to technological change but they can also actively influence how AI evolves to serve the public good, preserve cultural memory, and promote inclusive access to knowledge.

Understanding the Changing Behaviour of Users

While institutions evolve internally, users are also changing how they interact with information. Libraries have spent at least last two decades navigating the age of Google, continuously rethinking

their role in society.[3] Today, that challenge has become even more complex. Users across all generations are increasingly shifting from traditional library materials to generic online AI systems for both personal and professional purposes. As a result, their expectations around how they search for, access, reuse, and interact with information have changed dramatically, and in a remarkably short period of time.

Today's users expect intuitive search interfaces, personalised recommendations, and seamless access across devices. AI4LAM recognises this shift and emphasises **user-oriented services** that respond to evolving expectations. AI tools such as **chatbots, virtual assistants, recommendation systems, and analytical platforms** are being deployed to enhance user engagement. These tools not only facilitate access but also provide educational and cultural enrichment, supporting fields like science, tourism, and education.

Library systems enhanced by AI have the potential to fundamentally redefine not only how they serve users, but also who those users are.[4] We are no longer looking solely at individuals as the primary audience. The library users of the future increasingly include industries, large AI companies, and sectors such as healthcare, education, and government. These organisations seek access to curated, trustworthy data and advanced tools for research, innovation, and decision-making - positioning libraries as critical infrastructure in the broader digital ecosystem.[5]

GLAM Data: A Treasure Trove for AI Development

This shift in library users stems from the fact that libraries, along with all GLAM institutions, possess an extraordinary wealth of datasets that are ideally suited for the development and training of artificial intelligence systems.

These datasets encompass a wide variety of formats and sources, ranging from digitised cultural heritage materials such as scanned manuscripts, photographs, and audio-visual archives, to digitally born content including e-books, e-journals, institutional records, catalogues, databases, and curated digital exhibits. In addition, legal deposit organisations play a crucial role in collecting and preserving harvested web content - extensive big data collections that represent a growing body of research outputs and electronic resources.[6]

As we already know very well, metadata is essential for organising digital collections, enabling searchability, and enhancing user experience. For this reason, metadata represents one of the most

---

[3] Shuqing Li, Fusen Jiao, Yong Zhang, Xia Xu, Problems and Changes in Digital Libraries in the Age of Big Data From the Perspective of User Services, The Journal of Academic Librarianship, Volume 45, Issue 1, 2019, Pages 22-30, ISSN 0099-1333, https://doi.org/10.1016/j.acalib.2018.11.012

[4] Empowering Libraries in the AI Era: Ethical Engagement, Strategic Adoption, and Professional Development, https://www.ifla.org/events/empowering-libraries-in-the-ai-era-ethical-engagement-strategic-adoption-and-professional-development/, Last visited: 5 December 2025

[5] Javier de la Rosa et al., The Impact of Copyrighted Material on Large Language Models: A Norwegian Perspective, IIPC, 2025. The project "MIMIR" on copyrighted content, The project "MIMIR" on copyrighted content - NTNU, Last visited: 5 December 2025

[6] See IIPC, International Internet Preservation Consortium - IIPC, Last visited: 5 December 2025

impactful workflow use cases and is highly relevant to AI transformation strategies across all types of libraries. The creation, management, and provision of metadata through AI systems is central to the future of libraries.[7]

AI tools can automate metadata generation, improve the accuracy of Optical Character Recognition (OCR), and support multilingual access through translation systems. This automation not only increases efficiency but also allows institutions to scale their services and reach broader audiences. AI-enhanced metadata systems are particularly valuable for digitised and digitally born materials, harvested web content, catalogues, databases, and research repositories.

Importantly, GLAM data is not only valuable within the cultural heritage sector, it is increasingly recognised as a strategic asset for scientific research and private sector innovation. The structured and ethically curated nature of GLAM datasets makes them ideal for training AI models that require high-quality, diverse, and representative inputs.

Researchers in fields such as linguistics, history, computer vision, and digital humanities rely on GLAM data to validate hypotheses, develop new methodologies, and uncover previously inaccessible insights.[8] Meanwhile, private sector developers and start-ups are exploring partnerships with GLAM institutions to access these datasets for commercial AI applications, ranging from intelligent search engines and recommendation systems to educational platforms and virtual assistants.

What is more, metadata serves as the essential foundation for integrating artificial intelligence into cross-sectoral digital infrastructures. It provides the structured information that enables AI systems to understand, organise, and analyse cultural heritage content effectively. Without rich and accurate metadata, AI tools cannot perform tasks such as automated classification, semantic search, or content recommendation with precision. As such, metadata is not just a technical component of library catalogues and digital infrastructures, it is a strategic asset that underpins the success of AI-driven innovation.

In this context, GLAM data serves not merely as a repository of cultural memory but as a dynamic catalyst for innovation. By opening access to their collections and collaborating across sectors, GLAM institutions are helping to shape the future of AI, ensuring that it is informed by history, enriched by diversity, and guided by ethical stewardship.[9]

[7] Already integrated in workflows of some library systems

[8] See for ex.: Genatowski, Emily, Introduction to Artificial Intelligence Prompt Engineering, Introduction to Artificial Intelligence Prompt Engineering | DARIAH-Campus, Last visited: 5 December 2025

[9] See Open Heritage Statement, Open Heritage Statement, Last visited: 5 December 2025

Accelerating, Advancing, and Advocating: AI4LAM's Strategic Pillars

To strengthen the position of GLAM institutions within the emerging AI-infrastructure ecosystem, AI4LAM operates through three strategic pillars that define its mission and drive its transformative impact across the cultural heritage sector.

The first pillar centres on **accelerating the development, testing, evaluation, and integration of AI tools** into institutional workflows. This approach empowers GLAM organisations to adopt innovative solutions that enhance operational efficiency, improve accessibility, and respond dynamically to evolving user needs. By embedding AI into core processes, institutions can streamline tasks such as metadata creation, or user interaction, ultimately expanding their reach and relevance.

The second pillar focuses on **advancing professional skills** through comprehensive educational programs. These initiatives are designed to equip staff with the knowledge and competencies required to work effectively with AI technologies. From technical training in machine learning and data management to workshops on ethical AI use and digital transformation, these programs foster a culture of continuous learning and professional growth. This ensures that institutions not only adopt AI tools but also build internal capacity to sustain and evolve their use.

The third and most important pillar is dedicated to **advocating for systemic technological change** at global, regional, and national levels. AI4LAM actively engages with policymakers, funders, and institutional leaders to promote inclusive and responsible AI adoption. A key focus is on supporting smaller and under-resourced institutions by forming regional branches, sharing best practices, and facilitating access to tools and expertise. This advocacy work helps democratise AI innovation, ensuring that its benefits are distributed equitably across the GLAM landscape.

Together, these strategic pillars reinforce AI4LAM's commitment to sustainability, scalability, and meaningful impact. They provide a robust framework for guiding institutions through the complexities of AI integration while fostering collaboration, innovation, and ethical leadership in the digital age.

Community-Driven Innovation: The Heart of AI4LAM

AI4LAM's strength stems from its dynamic, community-driven model that unites a wide array of stakeholders across sectors. At the heart of this collaboration are GLAM representatives, including librarians, archivists, curators, researchers, and institutional managers, who bring deep domain expertise and a commitment to cultural preservation. Complementing them are university personnel such as professors, lecturers, students, and Open Science officers, who contribute academic insight and foster educational outreach as well as AI research. The community also includes IT professionals, ranging from AI developers and data scientists to investors, who provide the technical foundation and innovation needed to build scalable solutions.

This cross-sectoral synergy ensures that AI4LAM's initiatives are not only inclusive and forward-thinking but also grounded in practical, real-world needs. By working together, these diverse contributors help shape tools and strategies that reflect the complexity and richness of the GLAM ecosystem.

Through specialised working groups and regional chapters, AI4LAM facilitates localised engagement while maintaining a cohesive global vision - empowering institutions to collaborate, innovate, and lead in the responsible adoption of AI.

Real-World Applications: Internal and External Use of AI in GLAM

AI4LAM plays a pivotal role in advancing both internal and external applications of artificial intelligence within the GLAM sector. Its initiatives are designed to enhance operational efficiency, enrich user experiences, and ensure that AI technologies are ethically and effectively integrated into cultural heritage institutions.

Internally, AI4LAM supports a range of applications that streamline and strengthen institutional workflows. One key area is **metadata creation and enhancement**, where AI tools assist in generating descriptive and structured metadata, often at scale. This not only improves discoverability but also ensures consistency across collections.

AI is also being used for **workflow automation**, helping institutions manage repetitive tasks such as cataloguing, classification, and digital preservation processes more efficiently. Additionally, AI4LAM promotes the use of machine learning models to improve **Optical Character Recognition (OCR) accuracy**, particularly for historical documents and materials with complex layouts or degraded text, thereby unlocking access to previously unreadable content.

Externally, AI4LAM supports the development and deployment of tools that directly benefit researchers, educators, and the general public. For example, **chatbots and intelligent retrieval systems** are being implemented to provide more intuitive and conversational access to collections, enabling users to explore archives and databases through natural language queries.

AI is also being harnessed to create **research and educational tools** that support digital scholarship, such as automated transcription services, topic modelling, and semantic enrichment. Furthermore, AI-driven **segmentation and analysis of digital data**, including images, audio, and video, allows for deeper insights into cultural artefacts and supports innovative forms of storytelling and interpretation, when, for example, knowledge graphs are developed.

In addition to these applications, GLAM institutions are actively **testing commercially available AI tools**, evaluating their performance, and adapting them to meet the unique requirements of cultural heritage contexts. At the same time, many institutions are investing in the **development of proprietary AI solutions**, tailored specifically for real-time use within their environments. This dual approach ensures that AI technologies are not only functionally effective but also aligned with the ethical standards, preservation mandates, and user expectations that define the GLAM sector.

Through AI4LAM's support, GLAM institutions are becoming leaders in responsible AI adoption, shaping tools that respect cultural values, enhance access to knowledge, and foster inclusive engagement with our shared heritage.

Conclusion: A Future Built on Cooperation and Innovation

Within the CENL, a dedicated working group was established to support national libraries in embracing technological change.[10] This group maintains an active connection with the AI4LAM community and regularly presents its initiatives at the annual Fantastic Futures events.[11] The partnership between AI4LAM and CENL exemplifies how international collaboration can drive meaningful transformation across the cultural heritage sector. By embracing artificial intelligence, libraries and related institutions are not only modernising their operations - they are fundamentally redefining their role in society.

As AI continues to evolve, the GLAM sector must remain agile, ethical, and collaborative. AI4LAM's community-driven approach ensures that no institution is left behind and that the transformative power of AI benefits everyone, from librarians and researchers to students and casual users.

And indeed, the journey has only just begun.

[10] See Network Group: the AI4LAM Working Groups are open to participation: https://www.cenl.org/ai4lam-working-groups/, Last visited: 5 December 2025

[11] See Fantastic Futures 2025, AI Everywhere, All at Once, AI4LAM's annual conference, December 3 – 5, 2025, British Library, London, ai4lam - Fantastic Futures 2025, Last visited: 5 December 2025

Tan Lu:
Recognising Front Pages of Digitised Newspapers: Using Deep Learning and XAI

The Royal Library of Belgium (KBR), alongside other national libraries in Europe, plays an important role in archiving, enriching, and disseminating Belgium's historical and cultural heritage. Through various digitisation projects, KBR ensures that archival material is not only preserved but also made accessible to scholars, historians, and the wider public. One of the flagship initiatives, BelgicaPress, has already succeeded in digitising over four million newspaper pages. These digitised archives provide a rich resource for cultural, social, and historical research, enabling deeper insights into Belgium's history. However, such large-scale digitisation projects are not without challenges, particularly in terms of workflow optimisation and resource management. A relevant topic within KBR's digitisation workflows is the recognition of newspaper front pages.

Specifically, recognition of front pages is required for properly organising raw digitisation outputs for online publication. Currently, staff members in KBR often spend considerable amounts of time checking digitised outputs to confirm whether a page is indeed the first page of an edition. This manual process not only consumes time but also underutilises the expertise of library professionals, who could otherwise devote their skills to higher-level digitisation and curation tasks. Thus, automation of front-page recognition has become a practical need within KBR's digitisation pipeline.

In this context, the KBR Data Science Lab has initiated an effort to apply deep learning to automate front-page detection. The choice of deep learning is motivated by its success in a wide array of computer vision tasks, especially on image classification. Specifically, convolutional neural networks (CNNs) have proven particularly effective in recognising semantic patterns within images. For front-page recognition, CNNs can be trained to capture certain patterns that distinguish the first page of a newspaper from following pages. Such patterns may include title banners, headlines, and layout structures that are characteristic of front pages.

The project's methodology begins with curating a trial dataset of digitised newspaper images, including pages from 8 different newspapers (French, Dutch and German) over the course of a year. The dataset includes labelled examples of front pages and non-front pages, enabling supervised learning. A CNN-based model (Resnet) is then trained on this trial dataset to learn the distinctions between the two categories. In total, around 10K pages were used in the training-evaluation process, where about 2K pages were front pages. The evaluation of the trained model shows promising accuracy (0 mistake from 7 out of the 8 different newspapers, 1 mistake from 1 newspaper), suggesting that CNNs are indeed capable of supporting the automatic recognition of front pages within large newspaper collections.

Towards full-scale training, the project adopts a ResNet-50 model with approximately 25.6 million parameters. The training configuration emphasises computational efficiency, with input images scaled to 224x224 pixels. Using a dataset of 121 different newspaper titles over the course of one year, the training and evaluation split followed an 80/20 ratio, yielding roughly 130,000 pages, of

which around 30,000 were front pages. The model was trained using a basic cross-entropy loss without hyper parameter optimisation. After 50 training epochs, the system achieved similar-to-trial performance: a precision of 0.98, recall of 0.94, balanced accuracy of 0.97, and F1 score of 0.96. Specifically, for about 65 newspaper titles, the system made no mistakes at all; for around 10 titles, it made more than five mistakes, and for approximately eight titles, more than ten mistakes.

Yet, while accuracy is a critical measure, it is not sufficient on its own. For cultural heritage institutions like KBR, trust in automated processes is essential. Researchers and librarians must be able to understand and verify the rationale behind an AI model's decisions, particularly when those decisions affect large-scale digitisation workflows. This is where Explainable AI (XAI) becomes relevant.

The integration of XAI methods, particularly Layer-wise Relevance Propagation (LRP), provides a means of making the model's decisions transparent. LRP works by decomposing the output of a neural network back through its layers, attributing relevance scores to the input pixels. In practice, this means that one can visualise which regions of a newspaper page the model considered most important when classifying it as a front page. For example, the model might highlight the masthead, major headlines, or image placement as critical cues. Such visualisations enable librarians and researchers to assess whether the model is focusing on reasonable features, thereby validating the credibility of the classification.

With LRP, for each layer l of the CNN, it is possible to compute the relevance for all neurons, yielding a structured relevance vector. Such vectors allow us to trace how relevance flows through the network and to review the visual concepts it has learned at different levels of abstraction. They also enable the identification of 'prototypes', specific features or patterns that the model consistently uses in its decision-making. To support this interpretability, an auxiliary data analysis pipeline, following the line of work proposed in concept relevance propagation (CRP), was developed to systematically process these relevance vectors and make the internal operations of the CNN model more transparent.

The importance of explainability potentially goes beyond this particular project, as it creates a tangible paradigm for collaboration between human experts and AI systems. Librarians possess domain expertise that allows them to assess the historical context and structure of digitised outputs. By comparing their knowledge with the AI's focus, they can both confirm the model's reliability and identify cases where the AI may have learned spurious correlations. This synergy ensures that relevant digitisation workflows do not merely rely on black-box automation but integrate human judgment to maintain interpretability and trustworthiness.

The broader implication of this work is its investigation on the efficiency of practical workflows of cultural heritage digitisation. By potentially reducing the manual burden of front-page verification, AI can help to free staff to concentrate on high-level curatorial tasks. The combination of deep learning and XAI also represents a step toward a more intelligent, efficient, and trustworthy digitisation pipeline. It reflects a broader trend within cultural heritage institutions, where AI is increasingly being used not only for efficiency but also for enhancing research and discovery.

Looking forward, several avenues of development are envisaged. A first step is to deploy the trained front page recognition CNN model in operational workflows, Secondly, future iterations of the project may also explore hybrid approaches that combine visual and textual features for improved front-page recognition accuracy and reliability. In addition, continued refinement of XAI techniques will be necessary to handle more complex models while maintaining interpretability.

The experience gained through this project also contributes to the broader discourse on AI in cultural heritage. It highlights both the opportunities and the responsibilities of deploying AI in domains where accuracy, transparency, and trust are all in-focus. By documenting results and challenges, KBR contributes to a growing body of knowledge that can inform other institutions facing similar digitisation challenges.

Mona Lehtinen:
AI at the National Library of Finland

The National Library of Finland aims to be at the forefront of incorporating Artificial Intelligence (AI) solutions. With the open-source Annif tool, which we have been developing since 2017, we address the challenges of managing vast volumes of information and improving the discoverability of resources. Technologies such as Natural Language Processing (NLP) and Machine Learning (ML) play a crucial role in automating the subject indexing process, which traditionally relies on manual efforts by librarians. Currently we are also working on automatic extraction of bibliographic metadata using Large Language Models (LLMs). We are also actively participating in several project funding applications related to AI use in the GLAM (Galleries, Libraries, Archives, and Museums) sector, collaborating closely with our partners to advance these initiatives.

We do a lot of in-house software development, in which various AI assistant tools have become familiar to us. Furthermore, AI is also used in various other daily tasks, such as transferring information from old catalogues and card indexes of archive and manuscript materials to other information systems.

We are currently working on drafting our own AI Principles to guide our use of these technologies responsibly and ethically. We are also coordinating a project that relates to AI, EU legislation, and copyright.

Contributors

Taner Beyoğlu, General Director of the National Library of Türkiye

Carlo Blum, Deputy Director of the National Library of Luxembourg

Tamara Butigan Vučaj, Deputy Director of the National Library of Serbia, with Anđelka Zečević, Serbian Academy of Sciences and Arts

Anna Chulyan, Director of the National Library of Armenia

Paul Gooding, Professor of Library Studies and Digital Scholarship, University of Glasgow

Kalina Ivanova, Director of the National Library of Bulgaria, with Emilyan Pavlov, INSAIT

Yurii Kovtaniuk, National Library of Ukraine

Mona Lehtinen, National Library of Finland

Hannes Lowagie, National Library of Belgium

Tan Lu, National Library of Belgium

Trond Myklebust, Deputy Director General of the National Library of Norway

Martin Öövel, Director General of National Library of Estonia

Richard Ovenden OBE MA FRSA FSA FRHISTS FRS, Bodley's Librarian & the Helen Hamlyn Director of the University Libraries (the senior executive officer of the Bodleian Libraries), Head of Gardens, Libraries & Museums (GLAM)

Gilles Pécout, President of the National Library of France

Renata Sadunišvili, National Library of Lithuania, with Kristupas Morkūnas. Statistics Lithuania. State Data Agency and Ernestas Čižus. Statistics Lithuania. State Data Agency

Melissa Terras, MBE FREng, Professor of Digital Cultural Heritage within Design Informatics, Edinburgh College of Art, at the University of Edinburgh, UK

Ines Vodopivec, Secretary General at AI4LAM

Wilma van Wezenbeek, Director General of the National Library of the Netherlands, with Rosemarie van der Veen-Oei, Dennis Eijsten and Michel de Gruijter


CENL
Foundation Conference of European National Librarians

www.cenl.org