

**Statement of Dr. Carla Hayden  
Librarian of Congress  
Before the  
Committee on Rules and Administration  
United States Senate**

**“The Use of Artificial Intelligence at the Library of Congress,  
Government Publishing Office, and Smithsonian Institution”**

**January 24, 2024**

Madam Chairwoman, Ranking Member Fischer, and Members of the Committee, thank you for the invitation to appear before the Committee on Rules and Administration to discuss a topic that is currently of great interest in government and across the public archive sector – artificial intelligence (AI).

I am pleased to have the opportunity to further engage with my Legislative Branch colleagues on this topic and to update the Committee on the Library of Congress’ (Library) activities exploring AI to expand access to our collection, enhance services for users, and improve internal processes for increased efficiency. Within these efforts, the Library is taking a serious approach to cementing a slate of strong governance for AI, as well as taking advantage of our position as a leading national cultural heritage institution to foster collaboration amongst stakeholders and partners engaged in this quickly changing technology.

Becoming a more digitally enabled agency has been a key focus of my time as Librarian of Congress. To meet the demands of a modern knowledge institution, the Library has found that technology must be approached in a deliberate and strategic manner. It is why in 2019 the Library published a comprehensive Digital Strategy to guide the agency’s use of technology in an increasingly digital world. Building on this first major step, the Library fully integrated its Digital

Strategy into its 2024-2028 Strategic Plan published last October. This most recent Strategic Plan embraces the central idea that technology must be “baked into all that we do.”

### **Expanding Access to Library Collections, Enhancing Services, and Increasing Efficiency**

The Library relies on a vast array of high-tech tools to provide services to Congress and its constituents. As the official home of federal legislative information and a library with nearly 176 million collection items, the Library, like many other organizations in the government and the private sector, is exploring opportunities related to using AI. In fact, the Library has been using some applications of AI – a term that has origins dating back to the 1950s – for years. For example, the Library started implementing one type of AI over 10 years ago, taking advantage of machine learning in the form of Optical Character Recognition (OCR) to assist in the processing of documents and enable full text searching of digitized texts.

#### *Increasing Access to Collections*

Since 2018, the Library’s Digital Innovation Division, also known as “LC Labs,” has investigated the utility of AI and shared the results of research experiments conducted by the division with the public. Working with digital researchers we call “Innovators in Residence,” LC Labs has made its mark as the launchpad for innovating engaging uses of AI. A popular example is Citizen DJ, a music sampling application that allows users to remix and create music using free-to-use, non-rights restricted audio from Library collections. The project was incredibly popular when it launched during the pandemic and continues to bring younger audiences to the Library today.

Other exciting active AI use cases include experimenting with machine learning and OCR to help create metadata and machine-readable text for digitized documents. For example, OCR has

increased the discoverability of more than 20 million historic American newspaper pages through the Chronicling America project. Building on that technology, users can also search historic newspaper photos using an application LC Labs rolled out in 2020 called Newspaper Navigator. This app uses a visual content recognition model to extract historic images from American newsprint published between 1900 and 1963, vastly enhancing users' access to 1.56 million images with just the stroke of a few keys.

### *Enhancing Services*

Enhancing services to Congress and the public is also a major area of focus. Several Library service units are successfully demonstrating AI's use in bolstering the Library's information services. The National Library Service for the Blind and Print Disabled is experimenting with available machine learning models to synthesize and compress lengthy book descriptions into succinct and engaging content for the blind, visually impaired, and reading disabled communities. Similarly, our digital innovators have been working with the Copyright Office to test approaches for extracting data from historical copyright records. This project combines human skills and AI capabilities to make these handwritten or typed analog records more accessible and easier to search online. It is just one example of how a "Humans in the Loop" model for AI can result in a successful integration of technology with human skill – an important principle indeed because the knowledge and skills of well-trained human beings will always be critical to the work of the Library of Congress. And we released a Congress.gov API (application programming interface) in 2022 to make it easier for the public to access and use accurate, structured congressional legislative data. The Library is also exploring whether automated processes, including machine learning and natural language processing models, can expedite the

drafting, reviewing, and publishing of Congress.gov Congressional Research Service bill summaries.

### *Increasing Efficiency*

In addition to its potential with bill summaries, the Library's AI use case testing is focused on increasing efficiency and staff productivity where staff have indicated an acute need for automation. As an illustration, we are testing the ability of AI tools to extract geographic locations from within legislative text. The Library also continues its historical role of innovation related to library cataloging. When the Library first began using computers in the 1960s, it devised the MARC (Machine-Readable Cataloging) format that would evolve to become the standard used by most library computer systems. Today, we continue in this tradition of being an industry leader by testing how AI can benefit cataloging processes. For example, the Library is experimenting with the use of AI models to help make more efficient use of staff time in the cataloging of books by using AI-generated content to process bibliographic data. Currently, our use case tests related to improving efficiency are designed to analyze and measure the quality of outcomes. In the future, the Library will use this information to help our divisions better understand the capabilities of AI for streamlining workflows performed by their staff.

### **AI Governance at the Library**

While taking advantage of certain use cases and benefits of AI, the Library has also placed a significant focus on building a robust governance framework. The Library's existing technology governance and policies provide a strong, adaptable foundation to guide the use of emerging technologies like AI. Further, the Library's approach to implementing AI closely aligns with the practices of other federal institutions and is informed by the National Institute of Standards and

Technology (NIST) AI Risk Management Framework, which aims to improve the trustworthiness of AI applications. Agency policies and practices are also informed by helpful standards coming out of the Executive Branch, such as Office of Management and Budget recommendations and Executive Orders signed in 2019, 2020, and 2023. As the Library’s experiments with AI move into additional planning stages and implementation, we have designed our policies and governance frameworks so that they can be continuously updated to reflect the specific challenges and opportunities presented by these evolving tools.

### **Collaboration Across the Government and Private Sectors**

The Library believes responding to this fast-developing area of technology calls for collaboration across the private sector, government, and academia. Last year, the Copyright Office launched its initiative on copyright and artificial intelligence. It held several virtual public roundtables and webinars, and issued guidance to help the public understand the requirements for copyright protection for works containing AI-generated content. It also issued a public Notice of Inquiry, and received over 10,000 comments, which it is reviewing and plans to issue one or more reports this year.

For our work with external partners, the Library participates in the General Services Administration AI Community of Practice and is a leading member of the international AI4LAM (Artificial Intelligence for Libraries, Archives, and Museums) Secretariat, sharing the leadership of the Secretariat with the Smithsonian Institution in 2024. In recent years, we have convened experts and practitioners for a “Machine Learning + Libraries” summit and hosted a leadership workshop that brought the Smithsonian Institution, National Archives and Records Administration, and Virginia Tech together to discuss the future of AI for public archives. Such

events have resulted in a valuable exchange of ideas, and we look forward to continuing similar cooperative efforts in coming years as the potential for AI technology develops.

### **The Continuing Need for Human Expertise**

We know from our experience that adopting automating technologies, including AI, requires human intervention. The Library has long understood the benefits of automation for improving both the efficiency and quality of our work. The MARC record, for example, was developed to improve the management of our collections and data.

However, with limited exceptions, even the most advanced models to date do not offer the quality that would make them appropriate for full-scale adoption of AI for business processes. The Library will continue to require human, hands-on expertise. While we may see a shift in what staff do every day as a result of AI, at this time we do not anticipate artificial intelligence having an impact to reduce our staffing needs. On the contrary, the Library will need to hire more skilled professionals in data science, analytics, and other expertise in order to take full advantage of AI for the benefit of our users.

### **Conclusion**

Discovering the role AI has to play in enhancing services to Congress and our other users remains an ongoing effort. As the Library charts its course forward, we plan to draw on our history of technical innovations, our rigorous development of standards, and our stakeholders and partners to align possible uses of AI with values of transparency, accountability, and efficacy. We welcome Congress's input and feedback on this topic and appreciate the leadership of this Committee in this issue area.

As always, I thank you again for your continued interest and support for the Library of Congress as we continue our more than 220-year history as the steward of the national collection and a repository for a vast shared cultural heritage. I look forward to responding to any questions the Committee may have.