Statement of Judith Conklin
Chief Information Officer
Library of Congress

Before the
Committee on House Administration
United States House of Representatives

"Artificial Intelligence (AI): Innovations within the Legislative Branch"
January 30, 2024

Chairman Steil, Ranking Member Morelle, and Members of the Committee, thank you for this opportunity to appear before you today on behalf of the Library of Congress (Library) to share the work we are doing with artificial intelligence (AI).

As the Library's Chief Information Officer (CIO), I work directly for the Librarian of Congress, Dr. Carla Hayden. Dr. Hayden and I agree technology is baked into all facets of the Library of Congress. While AI has recently become a focus in government and industry, the Library began leveraging AI and machine learning technologies more than a decade ago.

The Library of Congress is uniquely positioned to explore AI. We serve Congress and the American people with a universal and enduring source of knowledge and creativity. We steward vast collections of the nation's cultural and intellectual heritage materials. We also function as the research arm of Congress through the Congressional Research Service (CRS) and Law Library, and we are the home of the U.S. Copyright Office. Therefore, we have an abundance of rich historical data, legislative, and Copyright records in our care.

We are fortunate to possess data necessary for thoughtful AI adoption. We have earned the trust of libraries, archives, and museums across the country who look to us for guidance and collaboration on standards for preservation, storage, and other aspects of information sciences. We recognize our responsibility to use AI with our extensive large-scale digital collections in a variety of formats, and the opportunities to engage the American public, as we travel this path of learning the impacts of AI.

Origins of AI at the Library

AI is defined in different ways by different people. At the Library of Congress, we trace the beginning of our work with AI to its incorporation into Optical Character Recognition (OCR) technology more than ten years ago. The Library has been a destination for researchers and scholars over the last two centuries who use our primary sources and unique collections. Initially, we saw the power of OCR technology to enhance the ability to search our collections. The most powerful example of this early AI in action can be seen in our Chronicling America initiative, the largest collection of digitized newspapers in the world, where we have been able to empower our users to search millions of pages of newspapers from across the country over a period of nearly two hundred years.

AI empowers the researchers, genealogists, historians, educators, creators, and other users our librarians assist to cast a wider net and journey through our vast collections faster than humans alone ever could. Here, we see a great opportunity and a new challenge. If we are going to fully examine the potential of AI, we realized we needed a strategic approach that considered a wider scope of the Library's rich data.

The Agency-Wide Approach to AI

The Library established the Digital Strategy Directorate in the Office of the Chief Information Officer (OCIO) in 2018 to guide the agency's path forward with digital innovation. Innovation is not possible without technology. We recognized digital work the Library pursues relies on its information technology (IT) foundation, which is built and sustained by OCIO. As the Library's centralized technology organization, OCIO is also responsible for providing the leadership, strategy, governance, and management for all agency technology. This makes OCIO a true strategic partner to the Library's other service units, including the Library Collections and Services Group (LCSG), CRS, and the U.S. Copyright Office. From this vantage point, we pursue AI experiments, strategic approaches, and governance across of the agency.

While our AI experiments across the Library are investigating different scenarios, the major finding from the experiments is consistent across the board: AI empowers, but it cannot replace, the people who use it.

The role of people at the Library of Congress is essential to the uses of AI. Every day, hundreds of catalogers provide records to thousands of libraries throughout the world to organize collections of human knowledge. They create detailed records for material in over four hundred languages to help researchers across the globe navigate our collections. Our legislative analysts help Congress and the American people understand complex issues and bill texts through historic and legal lenses. Our colleagues in the U.S. Copyright Office inspire creativity and bolster commerce by opening the doors to records of centuries of artistic works.

AI Experiments

Now more than ever, our historical, authentic data draws researchers, historians, and technologists from all over the world. We continue to use OCR to generate machine-readable text from digitized documents. OCIO and LCSG are in the second year of an experiment testing a range of machine learning models and architectures to generate data for bibliographic records in our catalog. Our findings to date reinforce that AI empowers people. Using multiple machine learning models, we found very high quality results for a few record fields, such as Titles, ISBNs, and very standardized fields. In some cases, our results were over 99% accurate. In other cases, such as the creation of subject headings and dates in the bibliographic record the results were far less accurate. For subject headings in our initial experiment, the AI experiments resulted in approximately 25% accuracy relative to a human cataloger.

For over five years, the Library has engaged with the public to experiment with our collections data. As much as possible, we make our data sets available to the public. Among other wonderful discoveries, this work has shown AI can identify and extract images from our largest digital

collection of historic newspapers. AI has also allowed users to hear and remix rights-free audio from our digital collections. Currently, our Sanborn Maps collection is being used by local communities to reconstruct and share historical neighborhoods in virtual reality.

Turning to legislative information, CRS has embarked on various AI related activities with OCIO. CRS established a CRS AI Working Group to examine developments, policy implications, and AI potential. OCIO and CRS are currently conducting an initial experiment to determine AI approaches that could be incorporated to assist with the creation of accurate, high-quality, and trustworthy legislative data and bill summaries.

As the U.S. Copyright Office continues its important work examining the legal and policy implications of AI for the copyright community, OCIO is collaborating with them to discern ways AI can help generate data from historic Copyright Registration Records to improve discoverability by researchers, economists, and creators.

Beyond understanding the AI technologies currently available, each of these experiments is designed to identify quality assessment activities, standards, and benchmarks, and work with expert staff to develop tools for continued assessment of AI-generated content in their areas of expertise and responsibility. As we incorporate AI technologies, we are working to use machine capabilities to their fullest extent, while striving never to lose sight of those who motivate and share in our work: the people.

Governance and the Future of AI at the Library

Recognizing our role as a trustworthy steward of information, the need for efficiency in serving our mission, and understanding we are a source for standards and best practices for libraries and research institutions around the world, we have intentionally pursued our AI work in tandem with the establishment of AI governance. As a federal cultural institution, we strive to meet standards across the U.S. government and serve as an active community member.

While small for a federal agency, the Library of Congress is the largest library in the world and a source for standards and best practices for the field of information sciences. It has always been our charge to ensure our staff have the skills and knowledge to make the Library the most efficient, effective agency it can be for the ever-increasing body of human knowledge.

Our AI endeavors, both internal and external, led the Library to carefully consider the role of digital innovation in the agency's long-term future. In October 2023, the Library released its FY24-FY28 strategic plan, *A Library for All.* While the Library once had a separate digital strategy, as a sign of our growing digital maturity, this new agency strategic plan embeds our digital strategy throughout. Library leadership has recognized our mission to engage, inspire, and inform Congress and the American people depends on our digital capacity and our approach to careful and thoughtful digital innovation. It has been agreed digital is "baked in the business".

Following the release of the Library's new strategic plan, an agency-wide Artificial Intelligence Working Group was established, co-led by the Principal Deputy Librarian of Congress and me in my role as CIO. With representation from units across the Library, the group will facilitate discussions of the most effective and responsible use of AI technology in service to our goals and

mission. Because the AI landscape is ever-evolving, this working group will allow the Library to align work across the institution.

We also continue to participate in conversations outside of the Library to ensure we have a clear view of the state of the field in this country and around the world. The Library participates in the GSA AI Community of Practice, engages in regular discussions regarding risks and risk mitigation with the National Institute of Standards and Technology (NIST), and is a leading member of the international AI4LAM (Artificial Intelligence for Libraries, Archives, and Museums). The Library monitors AI Executive Branch actions. We have provided regular reports to this committee on our progress with both AI experiments and governance. In addition to these activities, I serve as Chair of the Legislative Branch CIO Council. In the past year, the council has collaborated on AI initiatives being performed in our respective agencies. This year council will be establishing a Legislative Branch AI working group to further discuss our AI initiatives.

Conclusion

For more than two centuries, the Library of Congress has embraced the herculean task of keeping pace with human knowledge. Our growth as an institution is evidence that technology and innovation have created more, not less, information for us to collect, preserve, and present to Congress and the public. While computers and machines can perform remarkable tasks with data, we will always need people at the center of this work if we are to truly remain a source of authentic, enduring knowledge and creativity.

AI is empowering people at the Library of Congress in remarkable ways. In turn, our work with AI can empower people outside of the Library on their journeys to find rich and interesting information, engage with knowledge and experts in new ways, and indulge the very human traits of imagination, curiosity, and creativity. None of this would be possible without your continued support for technology at the Library. Thank you for this opportunity to share our work with you today.