

**THE USE OF ARTIFICIAL INTELLIGENCE
AT THE LIBRARY OF CONGRESS,
GOVERNMENT PUBLISHING OFFICE,
AND SMITHSONIAN INSTITUTION**

HEARING

BEFORE THE

**COMMITTEE ON RULES AND
ADMINISTRATION**

UNITED STATES SENATE

ONE HUNDRED EIGHTEENTH CONGRESS

SECOND SESSION

WEDNESDAY, JANUARY 24, 2024

Printed for the use of the Committee on Rules and Administration



Available on <http://www.govinfo.gov>

U.S. GOVERNMENT PUBLISHING OFFICE

WASHINGTON : 2024

COMMITTEE ON RULES AND ADMINISTRATION

SECOND SESSION

AMY KLOBUCHAR, Minnesota, *Chairwoman*

CHARLES E. SCHUMER, New York
MARK R. WARNER, Virginia
JEFF MERKLEY, Oregon
ALEX PADILLA, California
JON OSSOFF, Georgia
MICHAEL F. BENNET, Colorado
PETER WELCH, Vermont
LAPHONZA R. BUTLER, California

DEB FISCHER, Nebraska
MITCH McCONNELL, Kentucky
TED CRUZ, Texas
SHELLEY MOORE CAPITO, West Virginia
ROGER WICKER, Mississippi
CINDY HYDE-SMITH, Mississippi
BILL HAGERTY, Tennessee
KATIE BOYD BRITT, Alabama

ELIZABETH FARRAR, *Staff Director*
JACKIE BARBER, *Republican Staff Director*

C O N T E N T S

	Pages
OPENING STATEMENT OF:	
Hon. Amy Klobuchar, Chairwoman, a United States Senator from the State of Minnesota	1
Hon. Deb Fischer, a United States Senator from the State of Nebraska	3
Dr. Carla Hayden, Librarian of Congress, Library of Congress, Washington, DC	4
Hon. Hugh Nathaniel Halpern, Director, United States Government Publishing Office, Washington, DC	6
Meroë Park, Deputy Secretary and Chief Operating Officer, Smithsonian Institution, Washington, DC	8
PREPARED STATEMENT OF:	
Dr. Carla Hayden, Librarian of Congress, Library of Congress, Washington, DC	19
Hon. Hugh Nathaniel Halpern, Director, United States Government Publishing Office, Washington, DC	26
Meroë Park, Deputy Secretary and Chief Operating Officer, Smithsonian Institution, Washington, DC	33
FOR THE RECORD:	
Testimony for the Record Submitted by Marci Harris, POPVOX Foundation ...	35
QUESTIONS SUBMITTED FOR THE RECORD:	
Hon. Amy Klobuchar, Chairwoman, a United States Senator from the State of Minnesota to Dr. Carla Hayden, Librarian of Congress, Library of Congress, Washington, DC	40
Hon. Deb Fischer, a United States Senator from the State of Nebraska to Dr. Carla Hayden, Librarian of Congress, Library of Congress, Washington, DC	42
Hon. Shelley Moore Capito, a United States Senator from the State of West Virginia to Dr. Carla Hayden, Librarian of Congress, Library of Congress, Washington, DC	47
Hon. Amy Klobuchar, Chairwoman, a United States Senator from the State of Minnesota to Hon. Hugh Nathaniel Halpern, Director, United States Government Publishing Office, Washington, DC	52
Hon. Deb Fischer, a United States Senator from the State of Nebraska to Hon. Hugh Nathaniel Halpern, Director, United States Government Publishing Office, Washington, DC	54
Hon. Shelley Moore Capito, a United States Senator from the State of West Virginia to Hon. Hugh Nathaniel Halpern, Director, United States Government Publishing Office, Washington, DC	57
Hon. Amy Klobuchar, Chairwoman, a United States Senator from the State of Minnesota to Meroë Park, Deputy Secretary and Chief Operating Officer, Smithsonian Institution, Washington, DC	58
Hon. Deb Fischer, a United States Senator from the State of Nebraska to Meroë Park, Deputy Secretary and Chief Operating Officer, Smithsonian Institution, Washington, DC	59
Hon. Shelley Moore Capito, a United States Senator from the State of West Virginia to Meroë Park, Deputy Secretary and Chief Operating Officer, Smithsonian Institution, Washington, DC	62

**THE USE OF ARTIFICIAL INTELLIGENCE
AT THE LIBRARY OF CONGRESS,
GOVERNMENT PUBLISHING OFFICE,
AND SMITHSONIAN INSTITUTION**

WEDNESDAY, JANUARY 24, 2024

UNITED STATES SENATE
COMMITTEE ON RULES AND ADMINISTRATION
Washington, DC.

The Committee met, pursuant to notice, at 2:15 p.m., in Room 301, Russell Senate Office Building, Hon. Amy Klobuchar, Chairwoman of the Committee, presiding.

Present: Senators Klobuchar, Fischer, Padilla, Ossoff, Capito, Hagerty, and Britt.

**OPENING STATEMENT OF HONORABLE AMY KLOBUCHAR,
CHAIRWOMAN, A UNITED STATES SENATOR FROM
THE STATE OF MINNESOTA**

Chairwoman KLOBUCHAR. Calling this hearing to order. Good afternoon. I would like to thank Ranking Member Fischer, as always, and our colleagues for helping us put together this hearing. With us today is Dr. Carla Hayden, who is the Librarian of Congress. Thank you. We always enjoy having you here.

Also here is Hugh Halpern, the Director of the Government Publishing Office. Back again, thank you. As well as Meröe Park, thank you for being here, the Deputy Secretary and Chief Operating Officer of the Smithsonian. We understand that the Secretary is ill. Just—he will recover but was not able to make it, so we really appreciate it.

Today, we are going to talk about the very important topic of artificial intelligence and the agencies that play such a critical role in serving the American people. AI has the potential, as we know, to lead to incredible innovation by supercharging scientific research, improving access to information, and increasing productivity. But like any emerging technology, AI comes with significant risks and our laws need to be as sophisticated, as the potential threats are there to our own democracy.

Understanding these risks and benefits has been a major bipartisan focus of the Senate, with Senators Schumer, Rounds, Young, and Heinrich leading a series of nine forums since the fall, and a number of us on various committees working on proposals so that we are going to be ready when this hits.

I think, as you all saw from the reports of the robocalls in New Hampshire with the fake voice of the President or other things that

have also occurred to candidates on the Republican side of the aisle, this is not going to be one side or another. It is something that we as a Congress have to deal with to put some guardrails. That includes, of course, the work that goes on in these agencies.

At our hearing in September, all the witnesses agreed that AI poses risks to our elections, and we heard testimony on why we must work to put guardrails in place. That is why I am leading a bipartisan bill with Senators Hawley, Coons, Collins, and also joined with Senators Bennet and Ricketts, to prohibit fraudulent AI-generated content in our elections within the framework of the Constitution, for instance, allowing for satire and the like, and why we need to take other steps like disclaimers on ads that use AI so that the citizens of this country can actually believe that it is their own candidate or their opposing candidate who is speaking.

Another example of legislation going on. Senator Thune and I have joined together in introducing a bill to put in place common-sense safeguards for the highest risk, non-defense applications of AI, and improve transparency. I see that Senator Capito is here. She has also joined us on this important bill that is mostly coming out of the Commerce Committee.

Let us get to your stuff, AI and how it affects the work of the three agencies before us, the Library of Congress, Government Publishing Office, and the Smithsonian. While it is important that our three witnesses today speak to measures they are taking to safeguard against potential harms, they are also, I think it is important to note, using AI technology in their work to protect our country's greatest treasures, advance scientific research, and improve public access to information.

For example, the Library of Congress is testing emerging AI technology to expand how researchers can better use the resources they already have in their collections, which make up the largest library in the world, we note with much humbleness, such as a new AI tool that lets users instantly search through 1.56 million images from digitized historical newspapers to assist in archival research.

GPO is working to harness the efficiencies offered by AI to modernize how it makes information from all three branches of government more usable for the public since much of its work has expanded to digital publishing, hence their new name, printing to publishing.

As part of its work producing government documents, GPO is using AI to ensure quality control of items such as the material you use to print passports. It printed 22 million of them, as we learned at the last hearing last year.

Finally, at the Smithsonian, which is the world's largest museum, education, and research complex, researchers are exploring how to use AI to do things like tackle some of the most challenging problems in astrophysics, classify species of fish in the Amazon, and make collections more accessible, accurately identifying the contributions of women, I like this one, in historical texts, in which they were often identified in writing by only their husbands' names. That will be interesting what you discover with AI.

We must continue working to stay ahead of the curve, and I am committed to working in a bipartisan way with Senator Fischer so that our country can benefit, and your agencies can benefit from

the best of AI, while protecting against any threats. Thank you again, and I will turn it over to Senator Fischer.

**OPENING STATEMENT OF HONORABLE DEB FISCHER, A
UNITED STATES SENATOR FROM THE STATE OF NEBRASKA**

Senator FISCHER. Thank you, Chairwoman Klobuchar, for calling this hearing on The use of Artificial Intelligence at the Library of Congress, the Government Publishing Office, and the Smithsonian. I also want to thank our three witnesses who are here with us today.

I know we are all wishing Secretary Bunch a speedy recovery. I believe this is the first time the Committee has heard from the Library of Congress, the Government Publishing Office, and the Smithsonian at the same time, and Director Halpern, it is nice to see you back with us so soon.

The Library of Congress and the Smithsonian both safeguard a vast collection of our Nation's treasures, welcome our constituents to their beautiful buildings to learn and explore, and serve as crucial resources to the scientific and the academic communities.

The Government Publishing Office also performs vital functions for the American people as it produces, distributes, preserves, and publishes documents for all three branches of our government.

This is the Rules Committee's second hearing on the use of AI. As we discussed at the first hearing, there is no question that AI is transformative, and it is poised to evolve rapidly. While AI brings the possibility of creating efficiencies and competitive advantages across government, it also creates risks.

Understanding and weighing the benefits and risks of AI are necessary first steps before widely adopting the use of AI. Today, I look forward to hearing from our witnesses about how our legislative branch agencies are exploring potential applications of AI.

There are many important questions to ask, such as are these existing issues or specific challenges at our legislative branch agencies that AI could solve? Are there existing services that our agencies provide that AI could improve? Can AI provide new avenues for these agencies so that you can fulfill your missions?

I also look forward to hearing about the guardrails our agencies are putting in place to mitigate those risks of AI. We need to understand the cost of AI and AI's impacts on agencies' budgets.

Just as importantly, agencies need to understand the implications of AI use, especially the risks related to privacy. Understanding legislative branch agencies' use of AI is an important oversight question for this Committee. That said, we cannot lose sight of existing challenges facing the agencies that we oversee.

As the Library looks to hire a new director for the Congressional Research Service, and GPO seeks to maintain high quality staffing standards in the face of your retiring work force, it is important that this Committee supports their efforts to improve hiring and retention efforts.

Furthermore, as the Smithsonian endeavors to establish two new museums authorized by statute, we must maintain rigorous oversight of their efforts to tackle a deferred maintenance backlog, a chronic issue for that agency.

With that said, I look forward to a productive discussion about our legislative branch agencies' use of AI, as well as other issues that you face. Thank you, Madam Chair.

Chairwoman KLOBUCHAR. Well, thank you so much, Senator Fischer. I will introduce our witnesses. Our first witness, as I noted, is Librarian of Congress, Dr. Carla Hayden.

Dr. Hayden was sworn in in September 2016. She has done a tremendous job leading the Library. She previously served as the CEO of Enoch Pratt Free Library in Baltimore, and she received her undergraduate degree from Roosevelt University, and master's and Ph.D. from the University of Chicago, that I also attended.

Whenever I see Dr. Hayden, I always think about Barbara Mikulski, who was so proud that a Baltimore librarian got this job, and you have made her proud too. Thank you.

Our second witness is Director of the Government Publishing Office, Hugh Halpern, who has been in his position since his unanimous confirmation, it does not happen all the time, by the Senate in December 2019.

He previously worked on Capitol Hill for more than 30 years, including for former Speaker Paul Ryan. He received his bachelor's and master's degrees from American University, and his law degree from George Mason. Thank you, and we really enjoyed your attendance at your last hearing.

Our final witness is Meroë Park, who became Deputy Secretary and Chief Operating Officer of the Smithsonian in December 2019. Previously, she worked for 27 years at the CIA. Okay, I did not know that. When I read these bios, I learn new things all the time. You probably do not advertise that, but that is interesting. Including as executive director and chief operating officer, and she received her undergraduate degree from Georgetown.

Okay, now we are going to swear in the witnesses. Okay, if you raise your right hand, do you swear that the testimony you will give before the Committee shall be the truth, the whole truth, and nothing but the truth, so help you God?

Dr. HAYDEN. I do.

Mr. HALPERN. I do.

Ms. PARK. I do.

Chairwoman KLOBUCHAR. Thank you, and you can be seated. We will now proceed to five minute testimonies. We will start with Dr. Hayden.

**OPENING STATEMENT OF DOCTOR CARLA HAYDEN,
LIBRARIAN OF CONGRESS, LIBRARY OF CONGRESS,
WASHINGTON, DC**

Dr. HAYDEN. Thank you, Madam Chairwoman, Ranking Member Fischer, and Members of the Committee. Thank you for this invitation to appear before you to discuss a topic that is of great interest across the government and within the public archives sector, artificial intelligence.

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's activities exploring AI. At the Library, there are four areas of focus with AI, to expand access to our collections, to enhance services for users, to improve internal processes

for increased efficiency, and to implement a slate of strong governance for the use of AI.

Becoming a more digitally enabled agency has been a key focus of my tenure as Librarian. 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, we have now fully integrated our digital strategy into our 2024–2028 strategic plan, and the plan embraces a central idea, that technology must be, as our CIO, Judith Conklin says, baked into all we do. Since 2018, the Library's Digital Innovation Division, also affectionately known as LC Labs, has investigated AI and shared the results of its research experiments with the public.

Working with digital researchers we call Innovators in Residence, LC Labs has made its mark as a launch pad for innovating engaging uses of AI to expand access to our collections. A very 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 the Library's collections.

Other exciting AI use cases include experimenting with machine learning and optical character recognition, or OCR, to help manage 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 those historic newspapers for photos using an application LC Labs rolled out in 2020 called Newspaper Navigator.

Enhancing services to Congress and the public is also a major area of focus. In 2022, we released a Congress.gov API, application programming interface, to make it easier for the public to access and use accurate, structured Congressional Legislative data.

Additionally, several Library service units are successfully demonstrating AI's use in bolstering the Library's information services. 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 with AI capacity to make a written or handwritten analog record more accessible and easier to search online, and it is just one example of how humans in the loop can be the model for AI and successful integration of technology with human skill, because a very important principle is that well trained human beings will always be critical to the work of the Library of Congress.

Now, the Library's AI use cases have also been testing opportunities to increase efficiency and staff productivity. For example, our Cataloging Division is currently experimenting with AI to help staff more efficiently process bibliographic information and catalog records.

There are many opportunities with AI, but there also must be robust governance. The Library is in a strong position as our existing technology governance policies provide an adaptable foundation to use in guiding us with 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 NIST's AI Risk

Management Framework, which also aims to improve the trustworthiness of AI applications.

Responding to this fast developing area of technology calls for collaboration across the private sector and government, which is why the Library participates in the General Services Administration AI Community of Practice and is a leading member of the International Artificial Intelligence for Libraries, Archives and Museums Secretariat.

To conclude, discovering the role AI has to play in enhancing services to Congress and other users remains an ongoing effort. As the Library of Congress charts this course forward, we plan to draw on our history of technological innovations, our rigorous development of standards, and the input of our stakeholders and partners to align possible uses for the public.

We appreciate this opportunity, and we hope that you will join with us in making sure the values of transparency, accountability, and efficacy remain as we explore this opportunity.

[The prepared statement of Dr. Hayden was submitted for the record.]

Chairwoman KLOBUCHAR. Thank you. Mr. Halpern.

**OPENING STATEMENT OF HONORABLE HUGH NATHANIAL
HALPERN, DIRECTOR, UNITED STATES GOVERNMENT
PUBLISHING OFFICE, WASHINGTON, DC**

Mr. HALPERN. Thank you. Good afternoon, Chairwoman Klobuchar, Ranking Member Fischer, Senator Capito. It is always good to see you. I am pleased to appear before you today to share some of the potential uses of AI and related technologies at the Government Publishing Office.

GPO differs from the Library and the Smithsonian in that it is fundamentally a manufacturing operation. We publish, produce, and maintain materials for all three branches of government.

Our 1,600 craftspeople and professionals produce virtually all of Congress's documents, along with numerous other publications, and manufacture secure credential products like the United States passport. We also provide digital information either through our own trusted digital repository, *Govinfo.gov*, or by serving data to our partners like the Library where they use that data on sites like *Congress.gov*.

No matter what you call it, artificial intelligence, machine learning, or a large language model, GPO's operations are just as susceptible to disruption as any commercial firms, and that is not necessarily a bad thing.

My written testimony describes GPO's policy approach to this new generation of tools, so my statement today will focus on three potential applications for AI and related technologies in our day-to-day operations.

First, we believe these tools can improve our quality assurance process by automatically recognizing defects that a human inspector might miss. We already use a rudimentary form of this technology in the production of the current version of the United States passport. GPO uses equipment that optically scans the pages that will become the identity page in a personalized passport.

This equipment looks at each strip of three pages for variances that exceed the specifications for the material and rejects those pages that do not conform to the standard. AI technology has the potential to further refine this review, allowing the machines to learn what may constitute a natural variation that is within the specification and what is not.

This has the potential to reduce defect rates, lower waste, and free up our quality assurance team to focus on solving bigger quality problems as they arise.

Second, we see tremendous potential for supplementing our proofreading team. Proofreaders are very difficult to hire, and we need to free them from making routine, repetitive corrections and allow them to focus on more subtle issues that really require a human being to interpret.

One example is capitalization. GPO style says that we capitalize the letter “s” in the word “state” when we are referring to a political subdivision of the United States. Currently, we use computer scripts to perform global search and replace functions on documents to correct our most common errors, including that one. But those scripts are really blunt instruments.

For instance, they cannot tell the difference between the “State of Minnesota” and a “New York state of mind.” AI holds the promise of tools that understand context and know when text refers to one kind of state or the other. That will cut down on the need for our proofreaders to review and re-review, and then correct material that has already been run through our automated tools, and free them up to focus on more difficult contextual issues.

My final example comes from GPO’s public information mission. We have had great success in making congressionally mandated reports publicly available since the end of last year, with more than 180 of those reports available on GovInfo as of yesterday.

Most of those reports come to us as PDFs. While that is a good format to show how the printed document looks, it is not always the best format for viewing on a phone, a tablet, or even for folks with disabilities.

While GPO would like to get these reports in more flexible formats like XML, the agencies authoring reports are not always equipped to supply them that way. AI technologies hold the promise of allowing us to extract the information from a PDF, understand the document structure, and produce an alternative view that works on different kinds of devices, all without manual, time consuming work from our team.

These are just three examples where we see applications in GPO’s operations, and we are considering many pilots in the future. All of these are intended to act as a force multiplier for our team, allowing our folks to be more productive, and deliver a higher value for our customers and the taxpayer.

Madam Chairwoman, Ranking Member Fischer, thank you for the opportunity to testify before the Committee today. I look forward to any questions the Committee may have.

[The prepared statement of Mr. Halpern was submitted for the record.]

Chairwoman KLOBUCHAR. Thank you very much. Last but not least, Ms. Park, thank you.

**OPENING STATEMENT OF MEROË PARK, DEPUTY SECRETARY
AND CHIEF OPERATING OFFICER, SMITHSONIAN
INSTITUTION, WASHINGTON, DC**

Ms. PARK. Thank you. Thank you for the opportunity to present today. I will pass along your well wishes to the Secretary. Early in our time as an institution, we collected meteorological data from volunteers nationwide via telegraph.

Fast forward to a few years ago, when we helped create the first image of a black hole at the center of the Milky Way. In short, we have always embraced revolutionary technology, from the locomotive to the pocket watch, from the telegraph to the internet, and artificial intelligence is no different.

Throughout the institution, scientific uses of machine learning are actually quite common, applying it to the identification of exoplanets, tracing the evolutionary history of pollen, and conducting field research in conservation. In other words, artificial intelligence has been around the Smithsonian for almost a decade. What has changed is the widespread availability of generative AI.

These widely available applications of AI hold great promise, from making us more efficient and effective as an organization, to improving the visitor experience and extending our reach across the Nation. We are learning from our current use cases and are exploring new ones as this technology continues to evolve.

For example, the National Museum of African American History and Culture's Freedmen's Bureau Transcription Project, which is the Smithsonian's largest crowd sourcing initiative, is transcribing genealogical records of the formerly enslaved. In a recent trial run, AI tools reliably transcribed nearly two thirds of the test content, showing how AI has the ability to dramatically scale up work with large data sets.

Other examples, including, as the Chairwoman mentioned during her opening remarks, our data science labs' development of a model that can discover incorrect women's contributions mistakenly attributed to men in our collections.

The Smithsonian Astrophysical Observatories Astro AI Center has more than 50 AI projects planned or underway, creating excitement about our solar system and the universe itself.

Inside the Smithsonian, we recently announced the creation of an AI Community of Practice, which will help us test new capabilities across the institution, as well as build a framework of good governance to determine the best way to use AI effectively and responsibly.

People rely on our reputation as a trustworthy reservoir of knowledge, so we understand concerns about bias, ethics, safety, and accuracy in available AI outcomes, but there are also fantastic opportunities as well. We and other cultural institutions can collaborate with technology leaders to help improve AI tools for everyone.

We can also examine the many dimensions of AI from a scholarly perspective, convening experts to examine its ethical, social, and economic implications. The nation's 250th birthday in 2026 will be a pivotal moment for us to experiment with innovative AI tools, to enable collaboration with tech leaders in other government and cul-

tural organizations to make AI more reliable and trustworthy, to educate people about the benefits and the risks of AI.

Ultimately, AI is a tool. Used properly, it will allow the Smithsonian to expand our reach and our impact. Like any new technology, it comes with risks and the potential for unintended consequences. But by developing internal expertise and partnering with external experts, we can thoughtfully leverage AI, both for the Smithsonian and for the American public. Thank you.

[The prepared statement of Ms. Park was submitted for the record.]

Chairwoman KLOBUCHAR. Thank you very much. I am going to turn it over to Senator Fischer and—for her questions, and then Senator Padilla.

Senator FISCHER. Thank you, Madam Chairwoman. Thank you all once again for being here today. Can each of you share whether your respective agencies have an official policy that governs its use of AI? Dr. Hayden, do you have an official policy on that?

Dr. HAYDEN. Yes. The Library has a number of policies for security, IT security in general, and we have an AI working group that has developed guidelines.

We also follow the NIST's framework. We also are very cognizant and use the Executive Orders that have also laid that foundation. We are very, I would say zealous on making sure that we have those frameworks and guidelines.

Senator FISCHER. You mentioned in your oral testimony, it is in your written as well, about NIST and it aims to improve the trustworthiness of AI applications. Can you give me an example of how that would be used? I think a lot of us, when we think of AI, we do not get down to the nitty gritty.

We do not get down to the basics of what it really can do. We just think, oh, this is going to be so wonderful. It will be sorting data. I think this—I think some of your answers from all of you would be very helpful on that.

Dr. HAYDEN. One of the through lines for each of us is trustworthiness and with the Library of Congress, we take very seriously being stewards of historic information and current information so that we are a trusted source.

Working with the framework that includes looking at a new technology, you map it, you explore, and then you measure its usefulness and effectiveness. Then if it passes quite a few of these things, you would look at how you could implement it.

Being able to make sure and having these templates, we have questions that we pose when anything is being proposed. But with AI in particular, we ask what does it do? How can we make sure that humans are in the loop for their quality control? It is a very important part of looking at AI in particular and any newer technology.

Senator FISCHER. Thank you. Mr. Halpern, can you let me know if you—do you have a policy?

Mr. HALPERN. Absolutely. We have a directive in place since last month, since the end of 2023. One of the things that that created was our AI Governance Committee, and I believe that committee is going to have their first kickoff meeting on Monday, where they are going to start evaluating different technologies for some pilot

deployments in the near term so we can start testing some of these in our environment.

Senator FISCHER. You know, you mentioned in your testimony, a lot of—the three interesting examples you gave of how you are using AI right now.

How do you decide on a case-by-case basis when you are ready to test that on that project, whether you think it is safe at that point in time, or if you are just trying to gather information?

Mr. HALPERN. We already have a very robust system for evaluating software and hardware for deployment at GPO.

Much of what we do is mission critical to the Congress, so we need to make sure that anything we deploy does not interfere with our ability to deliver for our customers, whether that is the House or the Senate, or our State Department customer, or any other customer.

This Governance Committee is another layer on top of that. Looking at some of these new technologies, both to see if there are improvements to the way we do things that can deliver benefits to our customers and minimize the risk from those same technologies of introducing variables into our product. For instance—

Senator FISCHER. What does that mean, the risks of introducing variables—?

Mr. HALPERN. We have all read about the issues with ChatGPT hallucinating facts that are not there. The good news is GPO does not generate a lot of public information. We make other people's public information available.

Much of where we can leverage these technologies is really in the quality assurance world. But, for the same reason you need to double-check ChatGPT's output, we need to make sure that the optical system that is checking the quality of our passport identity pages is not introducing or missing quality problems.

From that perspective, we really need to deploy these systems in a limited way so that we can make sure they are doing what they are advertised to do so that we are not delivering a product to our customer that is of lesser quality than it should be.

Senator FISCHER. Can I have the Smithsonian response as well.

Ms. PARK. Of course.

Senator FISCHER. Thank you. Ms. Park.

Ms. PARK. Great. Thank you. I will address maybe two aspects of your question. Internally at the Smithsonian, we absolutely are focused on governance and also the sharing of best practices.

This community of practice that we have formed is really designed to do a number of things. One is to identify policies that we already have that need to be updated with what is a very quickly evolving technology. Also, things like protecting privacy.

You asked about specific examples of how does this actually play out in an organization. You have to make sure that if employees are using ChatGPT or some other generative AI tool, that we are thinking about how that data is being prepared and made available to AI to analyze.

There is a lot of thought going into how we go about doing that internally. Then externally, we also have a public mandate, as several of my colleagues have talked about here, as trusted sources for the American public.

We want to make sure that we are available to them, that we have our 177 years of experience with history, culture, science, education as a way to inform these tools, and also as a way to help evaluate the accuracy of information.

So that is still aspirational for us at this point, but we are thinking about both aspects of our internal governance, our internal opportunities for efficiency, as well as our ability to be a convener and to help in the more public realm.

Senator FISCHER. Thank you.

Chairwoman KLOBUCHAR. Okay, Senator Padilla. Thank you.

Senator PADILLA. Thank you, Madam Chair. As I have had a chance to share in this Committee before—before entering the Senate, I served as California’s Secretary of State, where a part of the portfolio was serving as a custodian of the state’s archives.

At the time, I made it a priority to dramatically increase public access to our state’s archives, and this involved a significant undertaking to digitize our archive materials to make them easily viewable, not just to the people of California, but to the world.

Question, first question is for Ms. Park. You noted that prioritizing digitization is one of the Smithsonian’s key considerations with respect to AI. How do you envision leveraging AI to increase public access to the Smithsonian?

Ms. PARK. Indeed. I will maybe make a couple of comments in response to your question. The first is that digitization is absolutely one of our priorities as an institution. We have been at it actually for a while. It is a behemoth task with over 150 million objects, and we are working hard to identify ways to prioritize that. We have a long ways to go in terms of having those materials digitized.

When you refer to your own experience in California, the first step, obviously, with AI is even getting the records available and readable and analyzable in a digital format. So that is job number one for us.

The second part, is we believe that, unlike a search engine that would be able to help the public discover certain items that we might have in our collection and learn about them, AI has the ability for the public to make connections among the documents and objects and items that we have in ways that we cannot do as humans.

That we can use generative AI, and we hope to make it available to the public as well, to be able to discover things that we did not even realize we had in our collections. I think there is some really exciting opportunities here, but as you noted, job one is getting more of our items digitized and readable in a way that AI can make use.

Senator PADILLA. I will let you in on a secret, maybe a good strategy. Whatever is most popular already, that is where to start.

Ms. PARK. The ruby slippers.

Senator PADILLA. Because that tends to be the most popular to everybody else. Just imagine the curating opportunities that are possible when you get the metadata and the keywords right. There are some good experts that will be helpful. Dr. Hayden, same question to you.

Dr. HAYDEN. The Library has been really at the forefront in the library field with looking at how digitized material could be made more discoverable, because if it is not discoverable, it is really not accessible.

One of the projects that I mentioned in my testimony was taking a 20 year project of digitizing and making available newspapers from all 50 states, but adding an AI instrument, a navigator. You can now do even more research into those newspapers. The photos that are in there, names, all types of things.

We have seen that that has made what was just basically making things, putting them into digital form from the vast amount of analog material we have in print material going back hundreds of years, just basically digitizing them. That is an accessible aspect.

But having tools that have people be able to get into them has proven very popular, especially when we allow the public to use material in a creative way, like our Citizen DJ program, where we put the material up there, music, all types of things, and people can use them.

These tools can be very helpful for having people make those connections between our various collections—we have collections in 470 languages, and being able to use AI tools to help—

Senator PADILLA. A follow-up question to both of you. Do you envision institution-wide or even specific initiatives to always be contained within your respective institutions, or are there opportunities for public-private partnerships that you are envisioning at this point? Dr. Hayden and then Ms. Park.

Dr. HAYDEN. Well, the Library of Congress and the Smithsonian are co-chairing a Secretariat in this coming year that will look at more ways that we can all collaborate and, of course, opportunities with the private sector when we are making sure there are guard-rails to them.

Senator PADILLA. Correct.

Ms. PARK. I would just add that absolutely, there are opportunities to work with all sorts of external groups. Higher education, companies, all are interested in this space.

We have formed several consortiums with different organizations to try to figure out if we can convene together to talk about these kinds of issues, to provide space for people to discuss the complexities of AI. Absolutely, there are ways for us to partner with others.

Senator PADILLA. Okay. Thank you very much.

Chairwoman KLOBUCHAR. Very good. Thank you. Senator Hagerty.

Senator HAGERTY. Thank you, Chair Klobuchar and Ranking Member Fischer. I appreciate you holding this hearing. Dr. Hayden, it is good to see you back here again—to all three of our witnesses. Dr. Hayden, I would like to start with you, though, if I might. We all know the Library of Congress houses the United States Copyright Office.

Dr. HAYDEN. Yes.

Senator HAGERTY. I think, as a result, you are in a unique position to ensure the proper implementation of AI. I think, we all know that AI offers many promising benefits, but with its increasing prevalence, it is not without drawbacks.

I would like to take this home, and you and I have talked about my affinity for the songwriting and music business in my home State of Tennessee. We are actually home to thousands of performing artists there in Tennessee, songwriters, artists, all of whom have the potential to be greatly impacted by artificial intelligence.

I am concerned that AI-generated deepfake recordings, voice clones, or other infringements on the works of copyright holders could cause serious damage to an artist's reputation and to their livelihood.

Earlier this month, Governor Bill Lee announced plans for legislation to protect the voices of performing artists by adding their voices and sound to Tennessee's Protection of Personal Rights law.

I am pleased that my home State of Tennessee is leading the Nation in providing legal protection for our artists and for our songwriters. My first question, Dr. Hayden, is given the growing pervasiveness of AI, how does the Copyright Office plan to protect the copyrighted works and uses of voices of performing artist at the federal level?

Dr. HAYDEN. The Copyright Office has been in the forefront of looking at the issues involved with machine generated works. In fact, they have developed a separate web page for people who are creators, giving guidance on that.

They have hosted a number of presentations and webinars, and our Register of Copyrights, Shira Perlmutter, has been in the last year in many, many international copyright sessions and meetings. They just issued a rule to get information, and they had over 10,000 responses.

They are going to be issuing more information about what they are hearing from stakeholders about how that can really be used, but also be protecting.

Senator HAGERTY. Can I get into—let us drill into this a little bit more in terms of information you are getting back from stakeholders. In August, the Copyright Office issued a notice of inquiry on copyright and artificial intelligence. The notice sought information and comments on a number of issues, including the use of copyrighted works to train AI models, the appropriate levels of transparency and disclosure that are required with respect to the use of copyrighted works, the legal status of AI-generated outputs, and the appropriate treatment of AI-generated outputs that mimic the personal attributes of human artist.

I am sure that many Tennessee songwriters and artists submitted information to the Copyright Office urging the implementation of several things. One, the requirements for obtaining appropriate licenses or authorization to ingest or otherwise use materials that are copyrighted, or that implicate a person's rights of publicity or privacy.

Second, requiring adequate recordkeeping and auditing. Third, ensuring that appropriate transparency is maintained by AI companies. Dr. Hayden, are these remedies that I just listed, are those the types that the Copyright Office is considering?

Dr. HAYDEN. Well, the office is currently reviewing those comments, over 10,000, and they cover the full range. What they are saying, and I have the information here, that they are really look-

ing at what could be infringement, the treatment of these types of AI-generated outputs, and currently they are preparing that report. I cannot comment further than that but they will—

Senator HAGERTY. I hope they will take into account the three points that I have raised that are certainly of interest to our industry in Tennessee. Rights to the voices and likenesses of performers are generally protected by state law rather than copyright law.

That includes laws providing a right of publicity, privacy protections, and remedies for misappropriation. Given this, in your opinion, Dr. Hayden, are statutory changes needed to adequately protect the name, the image, the likeness of performing artists?

Dr. HAYDEN. Our Register of Copyrights, Shira Perlmutter, is very involved with that. She will be working with Congress on developing any potential legislation or any other things that would impact the copyright law as it stands. Right now, it is producing works by a human being. That is the standard.

Senator HAGERTY. I certainly hope you will view my office as a resource and the people of Tennessee as a resource here. I think we can add significantly to the conversation. Thank you, Dr. Hayden, and I yield back my time. Thank you.

Chairwoman KLOBUCHAR. Very good. Thank you. Some of the risks of AI are already clear. I talked about what just happened with those robocalls and what we are seeing on—really in fake videos on both sides of the aisle. It is not just one side. I know the Republican Attorney General is investigating what happened in New Hampshire.

But we know that there is going to be other security issues, from our infrastructure, cyber-attacks. Could you talk about what your agencies have done to guard against cybersecurity threats to ensure that your systems remain strong in the face of potential AI threats? Each one of you could answer. Dr. Hayden.

Dr. HAYDEN. The Library of Congress, with the help of Congress, has been able to build a very, very strong cybersecurity foundation over a number of years.

In fact, our CIO, Judith Conklin, is a cybersecurity expert. So that has been a major part of what we have been able to bring to looking at the risks with this particular technology, AI. It fits into our governance—IT governance framework already, and we are making sure that we are not on the cutting edge with AI.

We are looking at our policies using the experiences that we have had. But there is a technology governance board within the Library, and AI now has a working group within that board to make sure that everything is aligned with that. We are very concerned about the trustworthiness of products that have AI.

Chairwoman KLOBUCHAR. Okay. Very good. Mr. Halpern.

Mr. HALPERN. Absolutely. Our team works very hard to make sure that our systems are as impregnable as possible. But we are also looking at some AI tools as helpful in those efforts.

One of the things that AI excels at is pattern matching. They may be able to see patterns in our logs that a human being might not readily identify. So that is actually a benefit. One of the other areas that is tangentially related is in the privacy sphere, searching materials produced by others for personally identifiable information.

Just like many, many other organizations, we have had issues in the past where it has been hard for us to catch every incidence of PII in our materials that we make public. But again, these AI based tools that can recognize whether it is a social security number or address or another piece of PII in the Federal Register, the Congressional Record, or anything else we publish, that is an additional layer of security that these tools can help us with, so that we do not inadvertently put something out in the public domain that is not supposed to be there.

Chairwoman KLOBUCHAR. Very good. Ms. Park.

Ms. PARK. We take our cybersecurity responsibilities very seriously, both in terms of our own internal practices, but also in terms of our public domain. We do comply with NIST standards when it comes to our cyber profile.

Our CIO is very much checking constantly to make sure that we are following those standards. Our Inspector General is also there to regularly check on our progress. The other thing we do is we regularly report and, participate in federal forums where you share information about the latest threats that might be coming through various channels, including, the emergence of AI and how it is being used.

For one, for example, phishing has become much more sophisticated now with the emergence of AI and the ability to create much more easily emails and other things that appear to be real. There are—so we are stepping up our testing of our own work force in their terms of their ability to recognize these things. There is a lot of work going on there.

Finally, I would mention there is an internal group led by our CIO right now that is really trying to develop or evolve our policy structure to make sure that we are prepared for whatever threats that might be coming from emerging technologies, including AI.

Chairwoman KLOBUCHAR. Are you all taking part in cybersecurity training? Do you think that will be important as we see this changing world, both opportunities and risks, with your staff?

Dr. HAYDEN. Definitely—

Chairwoman KLOBUCHAR. Is your staff doing—okay. Yes.

Mr. HALPERN. Absolutely. We are evaluating vendors now for some specific AI related training so that folks understand these tools, and both the good and the bad that they can bring.

Ms. PARK. We have an annual training requirement for all staff who have access to our network.

Chairwoman KLOBUCHAR. Okay. Dr. Hayden, Senator Hagerty talked to you a bit about copyrights, voices. When you appeared before the Committee last year, we discussed the Copyright Office's new initiative to examine copyright law and policy issues raised by AI. What are the lessons learned by the Copyright Office over the past year? Have they continued to see increases in the amount of copyright applications for AI-generated content?

Dr. HAYDEN. Yes, they have. They have actually had a number of cases. There have been, as I said, about 100 works that have had AI-generated and human authorship, and they have been registered by the office.

They are also taking a leadership role internationally when looking at what is the proportion of the AI-generated content and human, as well. They have definitely been on the forefront of looking at that. But right now, they have rejected several applications because of the amount of AI authorship.

Chairwoman KLOBUCHAR. One last question of you, and then I will have just one more. Your written testimony highlights the work being done to test potential AI tools for the National Library Service for the Blind and Print Disabled, including improving the accessibility of book descriptions.

We have talked about this way back, some of the work we have done with the blind. What guardrails has the Library put in place to ensure the content being produced by these tools is accurate for its users?

Dr. HAYDEN. That is part of the LC Labs, and working directly with a service unit, NLS, National Library Service for the Blind and Print Disabled. They have had quite a bit of success with the e-reader rollout.

They are looking at how AI can be used to help this population. There are some exciting opportunities, but they are working hand in hand to make sure that what is being presented to that community is trustworthy.

You can see that that is the main theme, I think for all of us, that we want to use this tool for a number of reasons, for efficiency, effectiveness, but also we have to make sure that there is that oversight of use.

Chairwoman KLOBUCHAR. Okay. Last, Ms. Park, you talked about the assembling the group of experts from across the Smithsonian to identify best practices, opportunities. Could you talk about what next steps are after that, what the timetable is?

Ms. PARK. Yes. This group is charged with doing a number of things. The governance piece is an important one, reviewing policies, and also finding what is working within the units. We have multiple museums and research institutes.

Figuring out where we can do some things to scale rather than have individual units try to do their own thing where we can, for efficiency sake. I have asked for them to do their initial work and come back with some recommendations for what we might need to do in the short term.

What are some of the immediate stopgaps that need to be put in place, whether it is governance related or a best practice related issue that we want to share. Then I suspect there will be some longer term things that will be—take a little longer to address, whether it is systems related, business process, or applications. I expect it to be a bit of a phased approach.

Chairwoman KLOBUCHAR. My last question is, since Judy Garland's shoes came up, do you know where she was born, Judy Garland?

Ms. PARK. Oh, no, this is a quiz.

Chairwoman KLOBUCHAR. Yes. Well, Grand Rapids, Minnesota.

Ms. PARK. Yes, I was going to say—I was going to say Minnesota.

Chairwoman KLOBUCHAR. They just sort of solved the crime of the missing shoes that were there, which I would just like to note

for the record, finally. That they have been returned. It is a sordid story. Yes, but they are back. They are back.

In any case, thank you very much. Do you want to ask any other questions, Senator Fischer? We have a vote that has been called and thank you for the work that you are doing, continuing to do, leading your agencies.

I speak for myself, I know other Members of the Committee, we feel good about your stewardship of these very important agencies, and we look forward to working with you on the important work you are doing.

As we discuss opportunities, risks, I am sure we will have more discussions about this. The hearing record will remain open for one week, and we are adjourned. Thank you.

[Whereupon, at 3:12 p.m., the hearing was adjourned.]

APPENDIX MATERIAL SUBMITTED

**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.



THE HONORABLE HUGH NATHANIAL HALPERN
Director
United States Government Publishing Office

**Prepared Testimony:
The Use of Artificial Intelligence at
the Library of Congress, Smithsonian
Institution, and Government Publishing
Office**

**Committee on Rules and Administration
United States Senate**

January 24, 2024
2:15 PM
301 Russell Senate Office Building

U.S. GOVERNMENT PUBLISHING OFFICE

732 North Capitol Street, NW | Washington, DC 20401-0001
www.gpo.gov | facebook.com/usgpo | twitter.com/usgpo | instagram.com/usgpo



Chairwoman Klobuchar and Ranking Member Fischer thank you for inviting me back to the Committee to offer my perspective on the use of artificial intelligence and machine-learning applications (AI) at the Government Publishing Office (GPO).

I'm grateful to you both for the Committee's support and honored today to appear with two extraordinary public servants, Dr. Carla D. Hayden and Secretary Lonnie G. Bunch, III.

About GPO

GPO differs from my colleagues' institutions in its role and function. While the Library and Smithsonian, at their core, maintain, protect, and preserve essential collections and make them available to the world, GPO produces, publishes, and maintains information on behalf of all three branches of government. It is this production function that makes us different—our more than 1,600 craftspeople and professionals produce publications like the Congressional Record, the Federal Register, and a range of Federal information products including secure credentials and the United States passport, the most secure identification document in the world.

GPO also makes many of its publications available as tangible materials and digitally through its ISO-certified trusted digital repository, GovInfo. Because GPO maintains its unique role in the production process, we supply data to many of our partners, including the Library of Congress and its Congress.gov site. Similarly, we built much of the infrastructure used by the National Archives as it makes the Federal Register and the Code of Federal Regulations available.

A Framework for AI at GPO

No matter what you call it—artificial intelligence, machine learning, or large language models—GPO's current processes are just as susceptible to disruption from AI as those in any other billion-dollar enterprise. There is a role for AI in our manufacturing and production environments as well as our administrative and overhead operations. In general, GPO is looking to use AI as an opportunity to augment our team's capabilities, improve their performance, and allow them to focus on more difficult problems, rather than as an opportunity to reduce headcount.

With that said, one of GPO's critical challenges over the next several years will be the retirement eligibility of our workforce. As you know, nearly half of GPO's team will be eligible to retire by 2027. Through the selective application of these technologies, we can both relieve pressure on our existing team to perform repetitive basic tasks and at the same time make some of these roles more attractive to the next generation of teammates.

Working closely with this committee, the Joint Committee on Printing, and the Committee on House Administration, GPO has taken a deliberate approach to the potential represented by AI and similar technologies.

First, we created a use-case inventory similar in spirit and scope to those required of executive branch agencies and made that inventory public on our website. This provides a level of awareness, accountability, and transparency, ensuring that we fully think through these kinds of technologies before we deploy them.



Second, we developed an AI policy consistent with relevant statute as well as best practices such as those described in the March 2023 NIST AI Risk Management Framework. GPO's AI policy was issued this past October.

The policy stipulates that the use of AI at GPO must comport with our four Agency values—honesty, kindness, effectiveness, and inclusiveness—and enhance efficiency, increase productivity, and promote safety. It also requires that use of AI capabilities be disclosed to GPO employees, contractors, and clients and that requests for the use of new AI applications be submitted for review and approval to GPO's Technical Change Control Board (TCCB), which evaluates any new application or software that GPO teammates seek to use. The policy further clarifies that violations of GPO's AI policy must be reported and could be subject to possible disciplinary action.

The policy also calls for the creation of an AI Governance Committee at GPO comprised of key business unit managers as well as for the training of every GPO employee who may be entrusted with using AI capabilities. We believe this emphasis on training will prove critically important in the years ahead because we need to ensure such tools are used in a responsible and ethical manner.

Our AI Governance Committee is currently under development as is an overall AI Strategy for the Agency. I recently signed the charter for that committee, and it should be up and running within the next few weeks. We hope our overall AI Strategy can be ironed out within the next few months.

Examples of Potential Applications for Artificial Intelligence and Large Language Models at GPO

GPO is first and foremost a manufacturing enterprise. We produce a variety of documents for all three branches of government, including daily publications like the Congressional Record and Federal Register, committee reports, bills, and more specialized documents like White House programs, the Federal Budget, and the House and Senate manuals. We also manufacture the United States passport, the world's most advanced identity document.

Whether it's modernizing our manufacturing operations or automating our back-office functions, there are many applications for AI-enhanced tools to make our craftspeople and professionals more productive. We have a few systems that use rudimentary AI tools currently and can see many more applications as this software matures.

These are a few of the examples that we're considering now.

Administrative Functions

Acquisitions

GPO's Acquisitions Services team procures everything from computers, consulting services, and pens to the specialized papers, inks, and substrates that GPO needs to produce its products. We use Oracle's Public Sector Contract Lifecycle Management (CLM) system to manage the Agency's contracted procurements and are currently evaluating its AI features for a near-term pilot. This will hopefully increase the accuracy of GPO's



procurements and free up the Agency's valuable contract specialists to focus on more complex issues.

Employee Communications

GPO utilizes a newsletter tool developed by Axios known as Axios HQ for many of our routine communications with our teammates, including my weekly message. From the start, this platform included prompts to assist in writing entries in Axios' trademark Smart Brevity® style.

In a recent update, Axios added several AI-powered tools to assist in writing headlines based on the text of an item and automatically converting text to the Smart Brevity® style. I've personally experimented with these tools, particularly with respect to headlines. I've found that currently the tools are effective ways to see alternatives and I've chosen the generated headline on a few occasions. I'm sure that they will get better as the tools mature. However, our initial experience has shown that these AI-powered tools sometimes produce better results than we can come up with ourselves.

Cybersecurity

Many of GPO's cybersecurity tools have either released or are planning to incorporate AI-powered tools to recognize cyberattacks in real time. As talented as our cybersecurity team is, AI-powered tools would have the ability to view massive amounts of data and flag potential attacks to allow our team to proactively repel them while minimizing false alarms. We have already started evaluating many of these offerings to start pilot programs this year.

Process Automation

Robotic process automation (RPA) is a technology that is designed to automate routine, repetitive, and rule-based tasks, often in back-office functions. RPA software often simulates mouse movements and keystrokes to move data between systems or generate reports. The classic example is software that automatically opens a spreadsheet, copies certain data, pastes it into another spreadsheet, and saves that new spreadsheet in a particular location.

In an organization like GPO where we have multiple systems—including some legacy mainframe applications—adaptive RPA software can help us get important information to decision makers in time for it to be actionable.

As RPA tools incorporate AI-powered tools, they will be better able to handle the kinds of minor data differences that would cause errors in earlier generations of this software. Similarly, the use of AI to generate these kinds of processes may allow GPO's teammates to automate tasks themselves without the need for significant IT resources. This helps everyone to be more productive.



Manufacturing and Production

Quality Assurance Image Analysis

GPO is incredibly proud to produce the U.S. passport as it has since the 1930s. Known as the “Next Generation Passport,” the current iteration of that product has been in full production since April of 2022. One of the key improvements to that document is the adoption of the new polycarbonate identity page, a component manufactured by GPO.

As part of the manufacturing process, GPO utilizes technology which optically analyzes each of these identity pages and automatically rejects those that do not meet our quality specifications. Given that we manufactured nearly 22 million passports in FY 2023 and have an order for an additional 22 million in FY 2024, this kind of technology is critical for us to meet our production targets and our quality standards.

Similarly, with GPO’s new inkjet presses, we use a similar technology that allows the operator to see prints as they move through the machine. The machines are intelligent enough to easily recover from errors after the operator fixes the problem. We expect that this software will evolve to include even more error checking functions to allow operators to ensure high quality at an elevated production tempo.

This is the kind of traditional pattern matching where current AI technologies excel. As this technology improves, we hope to have even better results in both quality and productivity.

Making Proofreaders More Productive

GPO’s dedicated team of proofreaders and keyboard operators are the backbone of GPO’s in-plant operations. Without them, we wouldn’t be able to deliver final products for Congress or the Federal Register.

They are also difficult to hire. It takes a particular type of person with the dedication to perform this important but exacting, repetitive work, usually at night. In the past we had been able to recruit from the newspaper and publishing industries, but today there are fewer and fewer proofreaders employed by those industries. While GPO is trying to grow the next generation of proofreaders through its apprentice program, it is slow going.

Having more intelligent tools to supplement their work will help to make these teammates more productive and let them focus on more complicated problems.

For instance, currently we use simple “find and replace” scripts to address common capitalization issues. One element of GPO style is capitalizing the word “State” when referring to a political subdivision. However, these scripts can’t discern between the “State of Minnesota” and a “New York state of mind.” That requires our proofreaders to go back and make sure that the words are capitalized in context.

AI holds the potential of delivering more intelligent tools that would be able to understand that context, learn as our proofreaders correct copy according to natural changes in language, and reduce the routine work our proofreaders need to do to prepare a manuscript. This will free our team to focus on more complex problems, like formatting of bills or making sure that references are correct. That would be a major improvement for our teammates and our workflow.

Making Public Information More Accessible

GPO is proud of its important role in making government information available. Last year, GovInfo had its highest number of retrievals yet—more than 1.1 billion in FY 2023. That's a 32 percent increase from the prior year and a total of 10.8 billion retrievals since GPO started making government information available online in 1994.

That information needs to be more than just available—it needs to be findable and usable. AI technologies hold significant promise in delivering on those requirements.

Cataloging

AI technology and large language models have a demonstrated ability to summarize materials and discover relationships. The library community is already testing some of these tools for cataloging materials. These tools could be used for generating abstracts and other metadata to assist our librarians in cataloging government information materials. Having these tools do the preliminary work, while our librarians review and edit the metadata for quality, will allow us to be more productive with our current headcount.

Screening for PII

Considering the importance of protecting personally identifiable information (PII), we are always looking for new tools that can find this information and prevent its inadvertent disclosure. In the wake of last year's disclosure of PII as part of the January 6th Select Committee's report and related materials, GPO established a contract with a vendor that uses a machine-learning model to scan documents for PII, such as social security numbers. Our initial work with this vendor has been successful and is another important tool to guard against inadvertent disclosure of PII. We hope to adopt other AI-enabled tools to scan and redact PII where appropriate.

Providing Alternative Views of Government Information

GPO recently delivered on its statutory obligation to make congressionally mandated reports available via its GovInfo trusted digital repository. To date, we have more than 176 reports as part of this collection and almost all have been submitted in the PDF file format. While PDFs are an excellent method for document preservation and showing a document as it's intended to appear in print, other machine-readable formats display better on mobile devices and can provide data in formats that can be utilized in new applications.

While GPO will continue to encourage agencies to submit congressionally mandated reports in machine-readable formats in addition to PDF, we also intend to explore AI or machine-learning-enabled tools that may be able to extract the information from a PDF and display it in a more accessible format. This would have application in collections other than congressionally mandated reports as well and would benefit our patrons by providing the information in multiple formats.



Conclusion

Thank you Madam Chairwoman, Ranking Member Fischer, and Members of the Committee for the opportunity to testify about GPO's AI initiatives. We've covered a lot of ground discussing both the tools that are currently in use in our operations as well as some ideas we have for future applications.

GPO is committed to deploying AI applications responsibly and in a way that supports our team, not replaces them. We look forward to continuing our work with the Committee as these efforts continue.

I'm happy to answer any questions you may have.

**Senate Rules Committee Hearing on Artificial Intelligence
Wednesday, January 24, 2024**

For 177 years, the Smithsonian has been devoted to the increase and diffusion of knowledge. To achieve our mission, we have always embraced the creative use of technology. In the 1850s, we issued weather monitoring equipment to volunteers nationwide, collecting meteorological data via telegraph. A couple of years ago, we helped use computation from the Event Horizon Telescope's several thousand images to create the first image of a black hole at the center of the Milky Way. In short, we have always looked to the future. Artificial intelligence is no different.

Throughout the Institution, scientific applications of machine learning are common. Scientists at the Smithsonian Astrophysical Observatory, or SAO, use it to identify exoplanets. Smithsonian Tropical Research Institute scientists use it to measure the evolutionary history of pollen through fossils. Researchers at the National Zoo and Smithsonian Conservation Biology Institute use it for field research in conservation.

But there are so many more widespread applications of AI that hold great promise, whether it can make us more efficient and effective as an institution, improve the visitor experience, or augment the creativity of our curators and exhibition specialists. Many applications are currently being explored, from improving our websites' search engine functionality with AI-generated keywords to transcribing audio files.

Our Data Science Lab, created to address a massive increase in digital data, is developing a new AI model that can discover and correct instances in our collections where women's contributions were mistakenly attributed to men.

SAO has also launched AstroAI, a center that brings AI experts and scientists together to tackle the most exciting and challenging problems in astrophysics. AstroAI has more than 50 projects underway or being planned that apply various AI techniques.

Our Head of Digital Transformation, Becky Kobberod, is the Smithsonian's first leader devoted to developing a pan-institutional digital strategy. She will chair a formal AI Community of Practice, co-chaired by Ellen Stofan, our Under Secretary for Science and Research, and Deron Burba, our Chief Information Officer. With their leadership, we are determining the best way forward to effectively and responsibly use AI to enhance our work. They will help us build a framework of good governance to ensure the safe and responsible implementation of AI.

In a world filled with untrustworthy information, people rely on our reputation as a trustworthy reservoir of knowledge. Currently, concerns about bias, ethics, safety, and accuracy in available AI products make it necessary for us to proceed with caution. But as a trusted resource with vast expertise and experience, this is a terrific opportunity. We and other cultural institutions can collaborate with technology leaders to help improve AI tools, not only for our own use, but for everyone's.

Not only can we implement best practices for using AI at the Smithsonian and use our own resources to help create better versions of it; we can examine the many dimensions of AI from a scholarly perspective. The Smithsonian convenes conversations about some of the most consequential issues of our time. We can do the same for artificial intelligence, bringing experts together to examine the ethical, social, and economic implications of this technology.

Our strategic plan lays out some bold goals for the Smithsonian's future: to be more digital in the way we reach the American people, nimbler and more effective in our operations, elevate our scientific endeavors, expand our educational efforts, and be a more trusted source than ever. AI touches all of them.

The nation's 250th anniversary in 2026 will be a pivotal moment for us, not only as our strategic priorities begin to come to fruition, but as a time to experiment with innovative AI tools. It will allow us to see how AI can support our mission. It will enable collaboration with tech leaders and other government and cultural organizations to make AI more reliable and trustworthy. It will be a moment to educate about the benefits and risks of AI. And, most importantly, it will signal the further embrace of digitization to make our resources more available.

Ultimately, AI is a tool. Used properly, it will allow the Smithsonian to expand our reach and impact in science, art, history, and culture like never before. Like any new technology, it comes with risks and the potential for unintended consequences. But by leaning on internal and external experts, I am confident we can methodically and thoughtfully apply AI for all who benefit from the gifts the Smithsonian has to offer.

###



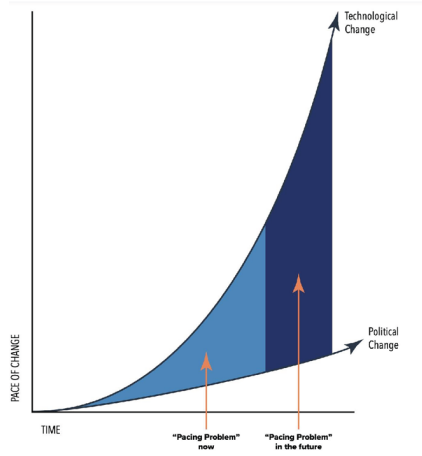
**Testimony for the Record of
Senate Rules and Administration Committee Hearing,
“The Use of Artificial Intelligence at the Library of Congress, Government Publishing
Office, and Smithsonian Institution”**

**Submitted by Marci Harris, POPVOX Foundation
January 24, 2024**

Chairwoman Klobuchar, Ranking Member Fischer, and esteemed members of the Committee on Rules and Administration:

Thank you for the opportunity to provide testimony to be included in the record for this important hearing, *The Use of Artificial Intelligence at the Library of Congress, Government Publishing Office, and Smithsonian Institution*.

My name is Marci Harris and I'm the co-founder and executive director of POPVOX Foundation, a non-profit with a mission to “inform and empower individuals and make government work better for everyone.” I'm also a proud former congressional staffer who loves this institution.



I have spent my career working to help governments address their “pacing problem,” a term coined by ASU Professor Gary Marchant to describe the effects of the exponential pace of technological innovation, paired with the lagging pace of government’s responsiveness through policy change and oversight.¹

ASU Professor Gary Marchant and his co-editors in *“The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight,”* described as “the pacing problem” - “the growing gap between the pace of science and technology and the lagging responsiveness of legal and ethical oversight society relies on to govern emerging technologies.”²

¹ <https://medium.com/g21c/congress-vs-the-pacing-problem-s-a887e3ca953f>
² Gary E. Marchant, Braden R. Allenby, Joseph R. Herkert, *“The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight — The Pacing Problem,”* Springer Dordrecht (2011)

In November 2022, OpenAI released for public use its generative artificial intelligence (GenAI)-empowered Large Language Model (LLM) ChatGPT. By the start of the following calendar year, the application had become the fastest growing consumer app of all time, with an estimated 100 million users in just two months.³ Although AI-enabled technologies, including predictive modeling, machine learning, natural language processing, computer vision, and speech recognition have been utilized by industry and government throughout the 2000s, GenAI is opening next level possibilities and putting powerful technological capabilities in the hands of anyone with a computer — without the need for advanced coding skills. This technology is revolutionary and everyone — technologists, academics, businesses and yes, even government — is starting at step one. In other words, GenAI is so new that government is *not yet* behind.

Congress' responsiveness to technological change has substantial downstream effects both in its internal day-to-day operational efficiency that drives service to constituents, and in its outward facing policymaking. Afterall, policymakers' familiarity with a technology affects their ability to properly scope policy and adequately govern it. An agile representational body able to transparently evolve with emerging technology, balancing risk with exploration, can unlock innovation within and beyond the walls of government. By engaging proactively, the Legislative branch can build its own confidence and resilience in the face of technological paradigm shifts.

Today's hearing is a formidable example of how the Legislative branch is keeping pace. Over the last year, I and my colleagues at POPVOX Foundation have studied international legislatures' response to GenAI in the context of the pacing problem. In our recent report, *Representative Bodies in the AI Era: Insights for Legislatures*, we make several recommendations the House should consider to minimize the institution's operational lag in adjusting to the new AI paradigm.⁴ I would like to highlight a few for the Committee to consider in the immediate term:

Initiate Early and Manage Timing Expectations

The Committee should focus its oversight efforts on encouraging AI-integration pilots and adoption in low-risk areas — across workflows both within Member and Committee offices, as well as internal business units across Legislative branch entities — to build a foundational understanding and capacity for these tools. By starting with non-critical functions, mistakes have limited consequences and important insights can be gained that, in turn, should be transparently shared with the public and stakeholders. If done with intention, the institution can emphasize its efforts to meet

³ <https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference>

⁴ <https://www.popvox.org/ai-vol1>

modern needs while maintaining a strong foundation built on human decision making and clear expectation-setting.

Invest in Upskilling

Familiarity with the technology – its benefits and its limitations – is essential for ensuring the institution is able to adjust to AI technologies' emergence. By investing in employee upskilling, an institution can foster responsible, measured experimentation and continue to monitor and address challenges that may arise. The Committee should partner with its peers on the Legislative Branch Appropriations Subcommittee to make resources available across Legislative branch entities for upskilling employees across their IT, cybersecurity, and other business units. Additionally, Legislative branch entities should continue to establish internal use-case guidance for employees wishing to augment their workflow with AI-enabled technologies and host upskilling training to ensure proper use of approved tools. This training is essential for minimizing risk and maintaining institutional awareness of AI-utilization across workflows.

Prioritize Data as a Strategic Resource

GenAI is revolutionizing how the public is accessing and digesting information. By standardizing legislative data and ensuring proper publication online, Congress can empower this emerging technology to capture Congress' original source documents, increasing the public's access to information about the work their legislature is undertaking.

The Legislative branch is the creator and steward of invaluable information in the form of data, utilized by Congress, federal agencies, and the public to ensure access to a historical record of legislative action. This data is a resource created, shared, and maintained through the collaboration of the House, Senate, internal congressional entities, and the Legislative branch agencies, including Library of Congress, the Government Publishing Office, the Congressional Budget Office, and the Government Accountability Office. Even though each of these entities plays an essential role in the life cycle of legislative data, a comprehensive data map wholistically examining the lifecycle of legislative data does not exist, nor does a comprehensive data management plan ensuring all entities manage data in a uniform, secure, and responsible approach. Although Congress has progressed thus far into the 21st century without these data-based assets, the emergence of GenAI has made them essential and the House and Senate need to lead the work of getting the Legislative branch's data assets in order.

Chief technology officers and chief information officers across the Legislative branch should be directed to engage in coordinating activities to begin creation of a

Legislative branch data map and begin creation of a comprehensive data management plan that ensures all entities are creating, storing, transferring and managing their data in uniform, secure ways.

Issue Agile and Transparent AI Guidelines

Entities across the Legislative branch should follow the precedent set by the House and Senate, and issue internal use-case guidelines for employees' utilization of GenAI-enabled tools. Clear, transparently posted guidance – that is supported by institution-hosted upskilling – will increase workforce confidence in exploring AI use cases while minimizing risks. However, it is important that entities revisit previously issued guidance on a regular basis to ensure it remains relevant and continues to address AI's ongoing evolution.

Customize AI Solutions

In systems integration, customized solutions are most effective when they are tailored to the specific needs and workflows of the adopters. The same is true for AI. A specialized approach for highly technical workflow integration increases usability of the platform while minimizing risk. For example, the Government Accountability Office was early in recognizing the efficiency gains possible through LLM technology, but also acknowledged the data access risks presented by commercially available models. As a result, the agency invested in the creation of a customized LLM, *Galileo*, that it launched for internal GAO employee use in 2023. The Committee should explore similar authority and funding flexibility across Legislative branch entities to allow for upfront investment and experimentation with customized solutions.

Engage in Global Collaboration

As explored in the *Representative Bodies in the AI Era* Report, AI is a global phenomenon. Legislatures and the entities that support them can benefit greatly from international collaboration and learning. By engaging with parliaments around the world that are pioneering the use of AI in legislative processes, legislatures can learn valuable lessons and avoid common pitfalls. A collaborative approach can accelerate learning and innovation, helping Congress to adopt AI in ways that are informed by a diverse range of experiences and insights. As such, the Committee should consider sending representatives to join discussions hosted by the International Parliamentary Union⁵ and participate in International Legislative Modernization Working Groups, such as the one hosted by POPVOX Foundation.⁶

⁵ <https://www.ipu.org/>

⁶ <https://www.popvox.org/modernization-working-group>



These recommendations are just a selection from a host of potential paths forward for the institution. But the important takeaway is: you should not stop.

I and the POPVOX Foundation team applaud the Committee's dedication to ensuring the First Branch is utilizing emerging technology to combat the pacing problem and keep Congress at the forefront of legislative modernization.

Thank you for the opportunity to provide testimony for the record.

POPVOX Foundation is a 501(c)3 nonprofit
with a mission "to inform and empower people and make government work better for everyone."
www.popvox.org info@popvox.org

Senate Committee on Rules and Administration
The Use of Artificial Intelligence at the Library of Congress,
Government Publishing Office, and Smithsonian Institution
January 24, 2024
Questions for the Record
The Honorable Dr. Carla Hayden

Chairwoman Klobuchar

At the hearing we discussed how artificial intelligence has the potential to impact your agencies' workforces in the years to come.

- 1. Can you expand on how you expect these tools to help your current employees to do their jobs more efficiently, while also creating demand for additional workers with skills relevant to this new technology?**

The Library of Congress (Library) has long understood the benefits of automation for improving the efficiency and the quality of our work. The Machine-Readable Cataloging (MARC) record was developed to improve the management of our collections and data. Optical Character Recognition (OCR) makes enormous collections of documents full-text searchable. As with previous automating technologies, Artificial Intelligence (AI) creates some efficiencies but also requires extensive human intervention, mediation, and management. As we have adopted new technologies that allow Library employees to perform their jobs more efficiently, staff skills have grown. We are relying on the foundation we have built for emerging technologies through our use of automated platforms, the creation of data structures, and our leadership and collaboration on digital standards and guidelines.

Currently, business units across the Library are exploring ways AI technologies might help manage and take advantage of the voluminous data in the Library's care. Thus far, we have intentionally focused our experimentation and research with AI on areas where staff have indicated an acute need for automation: assisting with the creation of summaries of legislative bills, generating data from historic records, proposing bibliographic record data, and providing access to non-text items such as images and audio. These activities, augmented by consultations with internal and external experts, have demonstrated how AI can enhance collections, operations, and services—but only by working with humans.

Many tasks at the Library require specialized knowledge and complex expertise, making efficiencies unlikely to be realized by commercial, off-the-shelf AI tools in the near term. Additionally, many uses of AI in our context require considerable effort to prepare data for automated processing. As tools emerge, we will continue to identify opportunities for efficiencies. We see an ongoing need for staff expertise and influence as these technologies are implemented.

We are working to identify the expertise workers need to respond to these new technologies. We believe these needs will span across IT infrastructure to data science, project and program

management, library and archival sciences, administrative support services, and a vast array of subject matter expertise pertinent to our collections and our communities of users. The Library continues to increase skills in data science and analytics, for example, and the Office of the Chief Information Officer (OCIO) Cloud Program Office, recently funded by congressional appropriation, is improving our capacity for managing big data. As we plan and experiment with AI, we are paying careful attention to the need for staff skills throughout the agency to manage and maintain these new technologies.

In response to a question from Senator Fischer at the hearing, you discussed the policies that your agency has in place or is developing to regulate its use of AI.

2. What specific information can you provide as to the anticipated timeline for your agency's next steps on these policies and related measures?

Through a series of regulations and implementing directives on IT governance, IT strategic planning, enterprise architecture, IT security, and IT investment, the Library has established a framework for managing information technology that includes evaluation of emerging technologies that should be strategic priorities for Library investment. This IT governance structure is headed by the Technology Strategy Board (TSB), co-chaired by the Principal Deputy Librarian of Congress and the Chief Information Officer (CIO) and including the heads of all Library service units, the Chief Financial Officer, General Counsel, and CIO experts. One of the chief duties of the TSB is to ensure that Library IT modernization efforts will meet strategic goals and deliver benefit to the institution and its users while addressing risks and challenges in a timely manner.

To engage with the AI's strategic potential for the Library, last year the TSB chartered an AI Working Group with representatives from business units across the Library and tasked the group with developing processes, policies, and governance specific to AI.

In November 2023, the AIWG published an initial planning framework for AI.¹ Currently, the AIWG is reviewing and describing current and future use cases for AI at the Library, preparing to publicly share an expanded use case inventory by June 2024. The group is also developing Library-wide processes, definitions, and principles, which will apply to the Library's custom, open-source, approaches with AI and to commercial off-the-shelf products incorporating AI features. These will be shared widely with the Library staff by August 2024. Finally, the group plans to have a formal governance model and draft AI Strategy for approval by the TSB by the December 2024. Policies and procedures resulting from these efforts will be incorporated into Library regulations and directives, as appropriate.

In the interim, in January 2024, the Library issued special guidance limiting the use of generative AI tools to produce materials for Congress or other government entities, for patrons or the public, or for internal audiences to pilots approved by service unit heads and coordinated with the AIWG (and TSB.) This guidance reflects the Library's fundamental role as a source of trustworthy and

¹ Available at: <https://blogs.loc.gov/thesignal/2023/11/introducing-the-lc-labs-artificial-intelligence-planning-framework/>.

authoritative data and the fact that, while AI's potential is immense, the tools still present risks with regard to accuracy, disinformation, bias, transparency, cybersecurity, and effectiveness, as well as privacy and rights issues. Even the most advanced AI models do not replace the expertise, experience, and insight that the Library's analysts, curators, librarians, and specialists can offer. The guidance limits the use of off-the-shelf generative AI products such as ChatGPT, BARD, DALL-E, MidJourney, and Stable Diffusion while the Library builds policy, guardrails, and quality controls for their use. This interim limitation on generative AI tools does not extend to AI that is embedded in regular office software used for word processing, graphics, etc.

The Library recognizes that there are some operational issues with AI that need to be addressed immediately, in contracts, for example. As a subgroup of the AIWG, the Library's General Counsel is convening a group to identify issues and make interim recommendations to address contracts, personnel, labor relations, and similar matters. As an initial step, for example, the Library is assessing whether Federal Acquisition Regulation clauses need to be adapted to be specific to AI and is soliciting feedback from its contractors on their use of generative AI tools to produce Library deliverables. This will inform future procurements and experiments and ensure compliance with AI best practices.

In the course of its work on internal policies and operations, the AIWG remains in touch with the USCO and its external policy work on the implications of AI under U.S. copyright law. This includes, for example, remaining sensitive to rights and privacy issues regarding the content that is used to train AI models.

Senator Fischer

1. Can you walk the Committee through how the Library is currently using artificial Intelligence? When did that start?

The Library currently implements AI where these technologies enhance our operations, and in smaller-scale, in-depth experiments and proofs of concept. Our use of AI began over a decade ago when we implemented OCR technology to process documents and enable full-text searching. OCR uses machine learning (ML) methods to process large amounts of text, and our adoption and continued use of this technology has helped to enable discovery in text documents such as the 20 million historic American newspapers in the Library's *Chronicling America Historic American Newspapers* database.

As AI technologies have matured, we have explored other ways to use them. The Library is currently using AI and ML to help Congress and the American public connect to Library materials in new ways, and keep our data trustworthy and secure. The Library's *By the People* crowdsourced transcription program uses "humans in the loop" methods to augment and validate OCR machine-readable text. The U.S. Copyright Office has an experiment to understand how effectively AI can assist in transcribing information from historical record books, adding to collections of digitized records, and enhancing the discoverability of online searchable information. CRS is investigating options to improve the availability of bill summaries and recently deployed new bill summary

workflow tools using Natural Language Processing (NLP) to assign bills to legislative analysts based on subject matter and help identify similar bills. OCIO is deploying AI and ML tools to discover cybersecurity threats and automate responses.

Because our mission includes both efficiencies and engagement, in 2020, the Library supported two proof-of-concept projects to demonstrate the capacity of AI to captivate and delight public audiences. These projects, produced by “Innovators in Residence” working with the OCIO and other units around the Library, used AI in music sampling and image searching applications. The Citizen DJ and Newspaper Navigator projects proved popular with users, inviting people to remix free-to-use sound collections in transformative ways and allowing users to browse over 1.56 million images extracted from historic newspapers. In addition, other visiting researchers experimenting with large-scale data analysis have used AI to search vast collections to find biblical quotations, identify objects in historic photographs, and classify records based on subject.

OCIO and partners across the Library are also conducting a series of experiments to explore AI and ML (described more fully below). This research focuses on addressing the ethical and practical challenges to adopting AI tools in libraries, archives, museums, and cultural memory organizations. It includes testing off-the-shelf tools, publishing state-of-the-field reports, and conducting in-depth experiments.

We are continuously capturing our uses of AI at the Library (whether fully implemented, proof-of-concept, or experimental) in an AI Use Case Inventory. We published an initial list on labs.log.gov, and plan to share an expanded version of this inventory.

2. Does the Library have plans to expand its use of artificial intelligence?

With a long history of applying new technologies and propagating standards for preservation, storage, and other aspects of information sciences, the Library began to experiment with AI applications years ago to identify appropriate use. These technologies are changing rapidly, and external users are eager to access our collections and data, including legislative information, sound recordings, historic newspapers, photographs, images, eBooks, and catalog records. Ethical and practical challenges must be overcome in adopting AI tools, particularly in our sector of cultural heritage organizations.

Our plans for expanding the use of AI rely on our research and experimentation before implementation approach. Our investigations have included broad possibilities and challenges, including a “Machine Learning + Libraries” summit, work with experts and practitioners, reports defining success and quality measures, modes of collaboration, and near- and long-term milestones. CRS has produced a number of objective, non-partisan products about the implications of AI available on [CRS.gov](https://crs.gov) and crsreports.congress.gov. The U.S. Copyright Office has closely monitored and researched policy and regulatory implications of developments in AI regarding creators of original expression and other rights-holders. Our expansion of AI must balance the huge potential for both increasing efficiency and creatively engaging the public, and the ethical and practical challenges of implementing these tools at scale. As a federal agency, we must carefully

consider the impact these tools will have on our staff and ensure that their expertise is integrated into any expansion of our use.

Expansion of AI technology use will be based on the OCIO LC Labs Planning Framework, which aligns closely with the National Institute of Standards and Technology (NIST) AI Risk Management Framework and the recommendations from OMB M-21-06: Regulations of Artificial Intelligence. Any recommended expansion will be considered by the Library's AI Working Group and the Technology Strategy Board. The planning framework, research, experiments, and staff oversight of AI implementation and expansion are designed not only to develop benchmarks for quality results and monitor these new technologies, but also consider the Library's mission and ecosystem in assessing those impacts before implementing at a larger scale.

3. Is the Library exploring ways AI can help solve specific issues the Library is facing? If so, what are those issues?

The Library has identified areas in which AI can help us accomplish our work. Our approach to AI is research and experimentation before implementation, and we currently have AI experiments that were identified and designed to address critical-need processes. These four experiments do not include all possible uses, but are priorities because of their cost-effective impact. The experiments are built to understand how AI technologies can build efficiencies that are not reachable without major staff increases. Regardless, as we continue to test various machine learning models, we have learned human intervention is required, and staff training for these roles is an essential part of planning and further implementation.

One current experiment by CRS and OCIO explores how AI approaches can expedite assigning geographic and organization subject terms to legislation; expedite drafting, reviewing; and publishing bill summaries; and determine baseline expectations of accuracy. Another experiment conducted by the U.S. Copyright Office and OCIO is working to extract data from page images of historic Copyright Record books. OCIO and the Library Collections and Services Group (LCSG) are looking at generating proposed data for bibliographic records and cataloging. The National Library Service for the Blind and Print Disabled (NLS) is experimenting with available models to synthesize and compress lengthy book descriptions into succinct and engaging content for patron discovery. These investigations address critical-need processes due to the scale of our collections and records, and our commitment to delivering high-quality services. Our research and experience in AI have demonstrated the importance of understanding how to best deploy AI tools in existing workflows, including establishing quality metrics and achieving the best outcomes from a combination of AI with our Library staff.

4. It seems like the Library is in the early stages of exploring AI and have not adapted its widespread use. Would you agree with that?

The Library has a methodical approach of experimentation with custom AI solutions before implementation at scale, but we have also implemented AI tools for widespread use in cybersecurity, digitization, improving workflows, and automation platforms. We have used OCR, which includes

AI, for over ten years. As the largest Library in the world, many other libraries, archives, and museums are interested in how we are using AI to solve common information science challenges and engage the public with very large collections of historical data. As the risks and the benefits are high for using AI at this scale, our AI experiments generate much more interest from the public than our use of automation platforms for supportive digitization, data storage, and business processes.

As a responsible steward of vast amounts of data and knowledge, we have learned AI tools are undertested or even untested with cultural heritage data and processes. Our research and experimentation led us to frameworks and processes aimed at developing criteria and workflows to evaluate AI tools and approaches prior to large-scale implementation. There are no one-size-fits all AI solutions to the Library's goal of providing access to and gaining insights from the vast amount of heterogeneous data collected over decades.

The top-line finding of the Library's first machine learning experiment was that individual tools perform sufficiently for well-defined, individual tasks, but that more exploration, evaluation, and frameworks are needed before broad application of ML technologies. The performance of different AI tools and approaches varies widely, and AI tools often underperform with the kind of historic, heterogenous, complex, and unstructured data the Library manages. For example, in our test of speech-to-text technologies, the tools were "highly inaccurate," likely due to the regional accents in the recordings and the poor quality of historical sound recordings.

We know care must be taken when applying AI tools to historical content that may contain outdated language, biases, and uneven representation. AI could amplify these aspects of our data in ways that could be incorrect, misleading, offensive to users, and tarnish the Library's reputation. For example, one research project on the post-Civil War Reconstruction era found terms and images from that era were represented incorrectly because of idiosyncrasies of collection metadata. Similarly, our experiment with catalog records has shown the lack of balanced subject terms affects the ability of AI models to generate subject terms correctly. For approximately 40,000 eBooks provided as training data, 30,000 unique subject terms were created—the vast majority of these applied in records only once. This resulted in less than 30% accuracy for AI generated subject terms.

Using AI technologies effectively involves understanding the underlying data, computer modeling, and human impacts. In some areas, implementing AI tools using standards and best practices such as those managed by the federal cybersecurity community are much more straight-forward. Thus far, our experiments have not demonstrated AI technology is able to meet our quality standards for widespread use in critical-need processes. Nonetheless, because these technologies are changing quickly, the baseline quality standards and benchmarks we are developing through research and experimentation, in combination with "humans in the loop" will allow us to adopt these technologies as they mature.

We will also continue to explore the unique possibilities of AI for engaging the public with the rich content we steward in ways that are unimaginable without AI. The possibilities for using AI to enable users to explore sounds, images, and texts extend well beyond our current descriptive practices, and we look for opportunities to make current practices more efficient. We aim to explore new practices to engage people in new ways without compromising our values of trustworthy and transparent provision of authentic information.

5. The Library submitted legislative requests aimed at making the Library more competitive in hiring and recruiting efforts. Can you provide some insight into these requests, and how these would help the Library improve its hiring practices and onboard high-quality candidates?

The Library continues to seek parity between its senior level positions and comparable positions in the executive branch. Specifically, the Library is proposing language to extend the authority in 5 U.S.C. 5753 and 5 U.S.C. 5754 to pay recruitment, relocation, and retention bonuses to individuals who accept senior level executive positions at the Library.

The aforementioned statutes currently authorize all agencies – including the Library – to pay recruitment, relocation, and retention bonuses to General Schedule (GS) employees. The statutes also authorize executive agencies – but not the Library – to request the Office of Personnel Management (OPM) to extend these bonuses to other categories of career employees. Accordingly, by regulation, OPM has authorized only executive agencies to pay recruitment, relocation, and retention bonuses to employees in career executive schedule (EX) positions, career senior executive service (SES) positions, and senior level (SL) positions.

The Library is thus at a distinct disadvantage compared to the entire executive branch, as well as the Government Accountability Office, when trying to recruit talent from top universities, the copyright community, the information technology industry, or other competitive fields. While other agencies can offer 25% higher pay and cover relocation costs to attract executive leadership, experienced managers, and high level technical expertise to public service, the Library cannot. Other agencies can also offer bonuses to retain talented, qualified leaders.

The proposed legislative request establishes parity for the Library, authorizing the Library to pay SL bonuses as set out in the existing statutes and requiring the Library to implement the authority through regulations consistent with the OPM regulations.

○ **How will these changes assist in the Library’s efforts to hire a new Congressional Research Director, and where are you in this process?**

The legislative request sought would allow the Library to pay recruitment, relocation, and retention bonuses to the candidate selected for hire, as applicable. Not having these hiring tools available to the Library could mean the difference between a highly qualified executive accepting or not accepting the position. The Vacancy Announcement for the CRS Director position opened on Friday, Feb. 2 and will close on Monday, Mar. 4. The interview panel will interview applicants and recommend finalists to the Librarian of Congress, who will then meet with finalists and make a selection after consulting with the Joint Committee on the Library. It is currently anticipated that the permanent position will be filled by late spring 2024.

6. Dr. Hayden, you appointed Dr. Robert Newlen as interim Director of the Congressional Research Service in June of last year. Can you tell us about Dr. Newlen's stewardship of the Congressional Research Service, has he implemented any changes?

Immediately upon Interim Director Newlen's appointment he began meeting with staff across the Service to conduct forums and consult with employees and Service leadership about how to best serve Congress. He gathered actionable feedback and designed a plan to implement changes, identifying quick wins, intermediate activities, and long-term goals. As a result of these efforts, the permanent Director, when hired, will be well-positioned to make long-term decisions about the organization and direction of the Service. In addition, Interim Director Newlen increased staff communication at all levels of the organization. He communicated clear expectations to the senior leadership team, and oversaw the redesign of regular senior leadership meetings to ensure they were effective and productive. In collaboration with the Chief Information Officer, Interim Director Newlen oversaw a wholesale assessment of CRS's short- and long-term technology needs. During his tenure, OCIO and CRS have partnered to release a new congressional request management system and TAP 2.0, a tool that will enable CRS to retire legacy systems and integrate with a new Bill Summary Workflow system for enhanced efficiencies, among other improvements. Interim Director Newlen continues to gather input and evaluate changes in an effort to ensure outstanding service to Congress consistent with CRS's core values.

o What will the onboarding process look like once a new director is chosen, will Dr. Newlen overlap with the new Director?

We anticipate that Interim Director Newlen will overlap with the new Director for some period of time. Interim Director Newlen is overseeing the development of a thoughtful, multi-faceted, and comprehensive onboarding strategy for the new Director, which will include introductions and meetings with key stakeholders, briefings on CRS's operations, and opportunities to hear from congressional clients and staff about how CRS most effectively execute its mission.

Senator Capito

1. How could AI modernize operations at each of your agencies and how could utilizing innovative technologies such as AI attract a leading-edge workforce?

The Library has long understood the benefits of automation for improving the efficiency and the quality of our work. The Machine Readable Cataloging (MARC) record, for example, was developed to improve the management of our collections and data. Well-structured APIs, such as the congress.gov API, sparks collaboration and innovation by providing access to large legislative data sets. A major draw of AI is its speed and efficiency. AI provides the ability to search across content, make assessments, and create information more quickly than humans alone ever could. Our research and experimentation strongly indicate AI will build efficiencies at the Library, especially with core processes for extremely large amounts of materials and those that are historically hard to digitize. It will be a force multiplier, but it also has the potential to generate an exponentially increasing amount

of materials submitted to the Library. Attracting and retaining staff with leading-edge skills is essential to this transformation, as there will be a continuing need for human involvement as these technologies are designed, implemented, and improved.

The Library's current experiments include the investigation of AI tools in core processes, such as proposing catalog record data, adding searchability for historic copyright records, and bill summarization. While we are hopeful, any AI implementations must adhere to defined principles and quality standards. Our next steps include creating benchmark data to test the effectiveness of new tools for specific tasks and formats and creating quality standards based on the outcomes of these tests. We predict we are still at least several years away from any wider application of trustworthy and secure AI tools—but only if we employ the time and care necessary to designing, testing and refining these tools within an organizational system that is able to support them.

Leveraging AI capabilities has the potential to improve efficiencies across Library operations. The future contracts workforce will employ automation, robotics, and artificial intelligence (enhanced search and analytics) to enable staff focus on refinement, presentation, and analysis of relevant information drawn from multiple dynamic public and private sources for better decision-making, risk identification and mitigation, and more efficient operations. Utilizing AI innovations will also enhance financial operations and capabilities to be performed around data analysis, data science (architecture), and independent cost analysis functions, as well as existing accounting, budget, and financial reporting and financial systems functions. AI tools will also improve the efficiency of personnel vetting through improved data retrieval and analysis.

Because our work with AI is at once marching in tandem with that of other federal agencies and organizations, and is highly specialized and in many ways unique, we often must participate in the creation of AI tools to meet our challenges, rather than waiting for industry to develop these tools. For example, most models are designed to work well with contemporary materials and data but are radically less effective with the Library's historic language, images, and artifacts. AI has been trained largely by materials on the contemporary internet, while our collections represent wider time periods, languages, and cultural perspectives. Current AI technologies are less adaptable to our needs, and subject matter experts remain vital to our operations. While this is resource intensive, our participation in open-source communities has opened doors to the sharing of best practices, collaborations, and attracting ledge-edge talent. Staff from OCIO LC Labs, co-leads a group discussing these issues with the Smithsonian, AI4LAMs (AI for Libraries, Archives, and Museums).

This leading-edge staff expertise is becoming ever more important to our use of AI. The Library of Congress is the largest library in the world and our products are used not only by over 15 million users a month, but also by many libraries, public and private, to deliver services. We attract seasoned experts in the information sciences, because of our mission and our scale but also because of our reach and the potential in serving Congress and the American people. As we incorporate new technologies, our approach to staffing is modeled after our continuous improvement methodology which includes partnerships with and co-ownership of technology applications. Our AI initiatives, projects, and experiments include technical experts and subject matter experts so that the breadth of staff that are engaged in implementation and continuous management are included from the start.

The Library continues to increase skills in data science and analytics, as well as other expertise across the agency to meet this need. The OCIO Cloud Program Office, funded recently by congressional appropriation, is helping us increase our capacity for managing data. The Library has hired and contracted for data scientists and data analysts in a number of areas including Copyright, the CFO's office, CRS, and LCSG. Staff participate in reading groups, internal conversations, and professional discussions to share information regarding AI. With that said, the Library does not currently have the capacity to realign resources to support larger AI initiatives moving forward.

2. Do you anticipate any procurement challenges in acquiring new technologies?

The Library's strong foundation of governance and policies guide our use of technology to meet the agency's mission, encouraging the adoption of tools and technology that will improve our ability to meet the information needs of Congress and the American people. These existing policies guide the use of emerging technologies, and the established AI Working Group under the direction of the Technology Strategy Board is charged with recommending amendments and the creation of new policies related to AI. The Library's approach to AI technologies includes research and experimentation, before implementation. This is critical to understand the range of benefits, risks, tradeoffs, costs, and standards before taking action.

As experiments move to additional planning and implementation, our governance and policies are updated to reflect the further application and use of AI. One specific regulation in use that is related to procurement is the Information Technology Investment Management Regulation (LCR 5-130). This regulation provides the framework for information technology investment governance and management at the Library, and it supports LCR 5-110, Information Technology Resource Management, which provides the foundation for an overall approach to information technology (IT) resource management at the Library.

Over three years ago, we created a new contract vehicle for experiments to support a methodical approach to defining, designing, and testing AI tools and the data used in the tools, both used and generated. As AI experiments are complex, multi-step processes, we have learned we must include specific contract language and require specific documentation from vendors to support our governance and mission.

Solicitations and contracts must state at a minimum: data that will be used, how the vendor can use data, data restrictions, and detailed documentation of data (training and generated) and technologies for accountability and traceability. Large procurement challenges include: the full disclosure by vendors of their models and training data, especially with commercial-off-the shelf products; verification of the data they collect and use from services; maintaining traceability through the multiple versions or releases, especially with commercial-off-the shelf products; and predicting long-term costs as these technologies change rapidly. As noted above in Question 2, the Library is also in the process of addressing our procurement processes and policies, including refining procurement clauses and terms as the Library continues to learn the benefits and risks of AI associated with larger scale use.

3. Dr. Hayden, with your agencies being so public facing, how could you use AI to enhance your exhibits and improve their accessibility to reach all Americans?

The Library has explored AI's use to enhance exhibits and discovery for all. Citizen DJ², which is our most popular, publicly available use of AI, allows people to access, listen to and remix rights-free audio from our digital collections, using AI to segment and visualize samples of audio collections. A version of Citizen DJ will be a part of the Library's new Visitor Experience. Newspaper Navigator is also publicly available and uses machine learning to identify and extract images from our largest digital collection of historic newspaper pages, with over 21 million newspaper pages, using the generative data for a keyword and visual similarity search. Our crowdsourcing program, *By the People*, uses OCR to generate a first draft of transcriptions for historical materials, which is then reviewed, supplemented, and edited by "humans in the loop", the crowdsourcing participants. This augmented and crowdsourced text is used in our digital collections to assist with discovery and availability in our IT systems and with assistive technologies.

We have several ongoing experiments to improve the discoverability of our collections and records online. One experiment is designed to understand the quality of generative data from Natural Language Processing for geographic and organization subject terms to allow users to discover legislation in new ways on congress.gov. Another is to extract data from page images of historic Copyright Record books, with the end goal of including these records as supplements to the public records search available on copyright.gov. Another is to synthesize and compress lengthy book descriptions into concise and engaging audio content for patron discovery on BARD and for other NLS services.

The Library of Congress is committed to digital accessibility. Digital Accessibility Standards (LCR 5-220) ensure the digital content, software, and websites the Library presents to the public and uses internally are accessible to individuals with disabilities. This regulation establishes standards for digital accessibility, responsibilities for ensuring compliance with the standards, and procedures for establishing exemptions from the standards.

As voice-enabled devices grow in popularity, AI tools will be used further to interpret and provide speech navigation, which is key to better serving our patrons through enhanced discovery. Currently, Congress.gov users can listen to bills that have at least one bill summary, bill texts, the Congressional record, committee reports, committee prints, and committee hearing transcripts. Users can also download the text as an MP3 file for later use. Increased digital accessibility leads to innovations for all. For example, well-structured front-end code and descriptive alt-tags used initially for assistive devices are used in emerging technologies. The captions that have now become ubiquitous in our virtual meetings were first requested by those with assistive devices and are now used in AI tools to analyze and translate text.

The list of services made available with AI technologies can both assist a range of disabilities and be used to enhance our exhibits and collections for all. This includes: creating audio summaries from large audio files, translating audio to text, generating audio from text, generating captions from audio, generating text and audio captions from visuals in video, breaking large video files into

² Available at: <https://citizen-dj.labs.loc.gov/>

sections with headings and summaries, generating summaries from eBooks, and generating audio and other content from geolocation markers with privacy controls. Our primary challenges at this time are the scale of our collections and the need to prioritize the use of our limited resources. Our current approach to AI of research and experimentation before implementation is used to balance the risks and the costs with using emerging technologies, but there is much that can be done to enhance access and services for all.

Senate Committee on Rules and Administration
The Use of Artificial Intelligence at the Library of Congress,
Government Publishing Office, and Smithsonian Institution
January 24, 2024
Questions for the Record
The Honorable Hugh Nathaniel Halpern

Chairwoman Klobuchar

At the hearing we discussed how artificial intelligence has the potential to impact your agencies' workforces in the years to come.

- Can you expand on how you expect these tools to help your current employees to do their jobs more efficiently, while also creating demand for additional workers with skills relevant to this new technology?

There are several areas in which we believe AI technologies can assist our workforce.

Quality assurance image analysis is one area where we think AI can improve our efficiency. We currently use a rudimentary form of this technology in our passport production operation to assist our staff reject passports that do not meet our customer's quality specifications and expect to extend similar technology to our press operations to check quality from our digital inkjet presses as well as new offset presses.

Another area of focus with AI is in assisting our professional proofreaders by absorbing some of the time-consuming, initial screening work they currently perform to enable them to focus their talents and skills on more complicated, higher value proofing and copy-editing challenges. We are currently working to hire additional proofreaders to assist with our workload and hope that AI technologies might soon be advanced enough to reduce some of the more repetitive parts of the job.

For our staff overseeing the Federal Depository Library Program, we also plan to evaluate AI tools that assist with the processing of bulk unstructured text into standard formatted documents with appropriate metadata. Those tools have the potential to jumpstart our ongoing efforts to digitize historical documents.

Procurement is another area where we hope AI can make a positive impact by enabling the automation of repetitive processes that can free both our acquisitions and customer services staff up to focus their energies on novel challenges and contracting issues while spending less time on rudimentary contract administration paperwork.

Fundamentally, we believe the broader deployment of AI technologies can help attract new professionals by reducing some of the frustrations some of our professionals currently face in the performance of the day-to-day aspects of their jobs, and by signifying that GPO is a place where new technologies are promoted, and new ideas are welcome. Fundamentally, we want our folks to have the best tools available to perform their best work.

In response to a question from Senator Fischer at the hearing, you discussed the policies that your agency has in place or is developing to regulate its use of AI.

- What specific information can you provide as to the anticipated timeline for your agency's next steps on these policies and related measures?

GPO recognizes that AI presents both significant opportunities and challenges and that we need to have standards and policies in place to prevent misuse. In October 2023, GPO issued GPO Directive 705.36, Artificial Intelligence Policy. This directive established "rules of the road" and provides best practices to GPO business units when acquiring, implementing, and utilizing AI products and services.

The policy requires that AI use at GPO must comply with applicable Federal laws and must align with GPO's values, promote efficiency, and promote the safety of our teammates and stakeholders and adopts the NIST AI framework and guidelines to ensure that GPO's policy reflects the best government standards. The directive also called for the creation of GPO's AI Governance Committee to oversee our approach to AI, determine appropriate uses, and identify potential risks.

The AI Governance Committee was established in January 2024. It will ensure that AI initiatives align with agency security standards, ethical standards, legal compliance, and organizational strategic goals.

The Committee held its first meeting on Monday, January 29, 2024, and decided to hire two AI techs, provide training agencywide, establish an AI lab, and pursue several pilot programs.

In addition, GPO also intends to have a draft overall AI Strategy document finalized before the end of this month that will provide a clear vision for AI integration and ensure alignment with organizational strategic goals. The AI strategy will focus on how the Agency can adopt and use AI services to achieve business goals and strategic objectives.

Senator Fischer

- Can you walk the Committee through how the Government Publishing Office is currently using artificial Intelligence? When did that start?
 - Does the Government Publishing Office have plans to expand its use of artificial intelligence?

Some of the AI tools we are using AI today are embedded in the updates to the software products we've used for years, such as the Microsoft 365 suite of products, and others are somewhat rudimentary. As a result, it's difficult to point to specific dates, but it's certainly true that any AI use at GPO is relatively new.

At the current time, GPO is using several applications that incorporate AI features to ensure efficiency, accuracy, and enhance operations. For example, our cloud-based cyberthreat detection engine uses machine learning and artificial intelligence to detect when a device is operating outside of its normal "known good" baseline. In addition, the tool we use to scan documents for instances of personally identifiable information (PII) before we post them on our GovInfo.gov trusted digital repository features machine learning capabilities. Our internal intranet site's search engine is also powered by AI cognitive features.

To make sure that the public is aware of the types of AI technologies and products we use, GPO publishes an AI use case inventory on our public website, which will be updated whenever we start using new products that incorporate AI technologies. That use case inventory can be found at <https://www.gpo.gov/explore-and-research/government-information/ai>.

In the future GPO believes AI technologies will present emerging challenges and opportunities that we must be prepared to meet. That view led us to create our AI Policy in October of 2023 and to the formulation of our AI Governance Committee just last month. Our AI Governance Committee recently held its first meeting on Monday, January 29, 2024, where we decided to hire two AI techs, provide training agencywide, establish an AI lab, and pursue several pilot programs during the current fiscal year.

Later this month, we hope to have a draft overall AI Strategy document finalized to provide a roadmap for how we intend to adopt and use AI services to achieve business goals and strategic objectives.

- Is the Government Publishing Office exploring ways AI can help solve specific issues the agency is facing?
 - If so, what are those issues?

Absolutely, as I mentioned at the hearing, GPO is first and foremost a manufacturing facility, so there are several potential process automations and AI-enabled improvements to analytical and diagnostic tools we hope to apply to our various production activities. For example, we'd very much like to be able to detect and remove defective passports from our production lines with less disruption to production flows. We would also like to improve our ability to

identify and remove misprints from our digital presses before they make it into bound volumes.

Other areas we are looking to incorporate AI technologies into our operations relate to our interest in reducing some of the more tedious, or repetitive, aspects of our professionals' workloads. We hope AI enable us to free them up to focus on more challenging and rewarding tasks and improve their ability to achieve their work goals.

One such area is in our proofing department where we hope AI-enabled tools can absorb some of the more time-consuming, initial screening work our highly skilled professional proofreaders must perform to permit them to focus their talents on higher value proofing and copy-editing challenges.

Another is within our Library Services and Content Management (LSCM) business unit where our staff perform cataloging and digitization work. We hope that AI tools might assist our LSCM staff with the processing of bulk unstructured text into standard formatted documents with appropriate metadata. These professionals are tasked with making incredibly voluminous amounts of public material available to the public, so the use of any tools that can help them reduce the amount of time needed to process individual publications is surely in the public interest.

- It seems like Government Publishing Office is in the early stages of exploring AI and have not adapted its widespread use. Would you agree with that?

Yes, very much so. It certainly feels as though we are the cusp of some significant technological changes and GPO's efforts over the past several months have been focused on making sure we the policies, governance structures, and personnel in place to help us reap benefits from the adoption of AI technologies while avoiding any pitfalls or serious missteps.

- The Government Publishing Office submitted legislative requests aimed at making the agency more competitive in hiring and recruiting efforts. Can you provide some insight into these requests, and how these would help the Government Publishing Office improve its hiring practices and onboard high-quality candidates?
 - Director Halpern, how would providing parity between GPO and Executive Branch agency policies help you recruit the next generation of employees?

With about half of our workforce becoming retirement eligible over the next four years, it is critical that GPO recruit its next generation of teammates. Retaining our most experienced teammates long enough to allow them to mentor this next generation is key to our success.

As a medium-sized legislative branch agency, GPO competes with larger executive branch agencies for talent. GPO's pay and benefits must be competitive with those in the executive branch for us to recruit successfully. Some of those barriers are statutory, such as the differences between the executive branch's SES and GPO's SLS programs.

One of the biggest disparities between SLS and SES benefit structures is in the maximum amount of leave our SLS employees can carry over each year versus their SES counterparts. GPO's SLS employees can only carry over 30 days of annual leave versus SES employees who can carry over 90 days. At a minimum, this makes it difficult for GPO to attract a talented SES-level manager who might otherwise wish to join our agency from the Executive Branch. It also makes us less competitive in hiring SLS/SES-level talent from the outside.

To improve our competitiveness as a potential employer, one of the legislative proposals have shared with the Committee would provide us with parity for the 15-20 SLS-level positions we employ at any one time in our 1,600+ person agency.

I am also seeking authority for a limited number of personnel appointments for confidential or policy-making roles at the Director's discretion. These appointments provide some additional flexibility in hiring individuals in the Director's Office, replicates similar authority for other agency heads, and protects against potential conflicts between the executive and legislative branches.

While virtually all of GPO's team is hired competitively using the procedures of the Office of Personnel Management (OPM), the GPO Director currently only has the authority to appoint three positions—the Inspector General, the Superintendent of Documents, and the Deputy Director—at his sole discretion. Any other appointment made outside of OPM's competitive procedures must be considered through the Schedule C process and is subject to approval of the White House, even though GPO is a legislative branch agency.

Whether the Director is looking to hire an executive assistant, a non-career Chief-of-Staff, or a particular specialist to focus on a special project, an officer of the legislative branch should not have to seek permission from an executive branch entity.

This proposal provides the Director of GPO with limited authority to appoint up to four individuals to confidential or policy-making positions. With this authority, my office would have the ability to staff special projects of the type that are needed to address key agency challenges like hiring and attracting new talent without OPM's cumbersome hiring procedures or interference from the executive branch.

Lastly, GPO is seeking to modernize the list of duties that those GPO employees detailed to Congress can perform. GPO currently details over 55 employees to House and Senate Committee and Leadership offices to help with the production of congressional publications. Under current law, written before the advent of computers and electronic publishing, those employees detailed from GPO can only perform work related to printing and binding of documents. Since most publication work is now digital, updating the definitions of the skillsets of GPO detailees would clarify the authority of GPO to both provide Congress with more relevant services and enable GPO to attract new talent to perform congressional work.

Senator Capito

- How could AI modernize operations at each of your agencies and how could utilizing innovative technologies such as AI attract a leading-edge workforce?

We think there are several areas where we believe we can leverage AI technology to improve efficiency, effectiveness, and the delivery of services. Quality Assurance Image Analysis is one area where we think AI can improve our efficiency both by helping our passport production staff identify and reject passports that do not meet our quality specifications much more quickly and by assisting our press and bindery staff's efforts to spot and remove misprints made by our newer presses.

We also believe that AI can assist our professional proofreaders by absorbing some of the time-consuming, initial screening work they currently perform to enable them to focus their talents and skills on more complicated, higher value proofing and copy-editing challenges. We are currently working to hire additional proofreaders to assist with our workload and hope that AI technologies might soon be advanced enough to reduce some of the more repetitive parts of the job.

For our Library Services and Content Management (LSCM) staff, who oversee the Federal Depository Library Program, we also plan to evaluate AI tools that might assist with the processing of bulk unstructured text into standard formatted documents with appropriate metadata. Those tools would improve our ability to catalog digital publications and to digitize historical documents.

Regarding our public-facing systems, we are exploring how an AI chatbot might be employed to communicate with external customers to provide details on services we offer and products we manufacture. Before we implement any AI technologies in this area, we would be sure to brief the Committee on any plans.

Other areas we can see AI technologies improving our operational efficiency include in our procurement processes, overall process automation efforts, and in the identification of active cybersecurity threats.

- Do you anticipate any procurement challenges in acquiring new technologies?

We don't currently see any AI-specific procurement challenges but recognize that the expanded use of new technologies requires a renewed vigilance to ensure that we are getting good value for our procurement dollars. We know that we will need to train our contracting officers to understand this emerging technology and incorporate subject matter experts into our procurement evaluations to make sure that the Agency's investments produce a good return on for the taxpayer.

Senate Committee on Rules and Administration
 The Use of Artificial Intelligence at the Library of Congress,
 Government Publishing Office, and Smithsonian Institution
 January 24, 2024
 Questions for the Record
Deputy Secretary Meroë Park

Chairwoman Klobuchar

At the hearing we discussed how artificial intelligence has the potential to impact your agencies' workforces in the years to come.

- **Can you expand on how you expect these tools to help your current employees to do their jobs more efficiently, while also creating demand for additional workers with skills relevant to this new technology?**

We recently launched an Artificial Intelligence Community of Practice at the Smithsonian (described more fully in response to the question below) to look at emerging tools related to AI and how these tools can bring efficiencies in various administrative domains such as contracting, finance, and human resources. Our goal is to identify a few key pilot projects to test promising opportunities and to automate steps in processes that will enable staff to do their work more efficiently and with higher satisfaction. We also expect many of our existing vendors to incorporate AI into the products that are used by our administrative teams that will also bring efficiencies. These AI enhancements will be evaluated and documented through established processes.

We do expect the demand for workers with AI expertise to increase as we expand the use of AI to accelerate our research, enhance our collections, and make our content more accessible to the American public. Establishing widespread use of AI for these purposes will require additional workers with expertise in these tools. We're currently addressing this need through partnerships and collaborations with other agencies and organizations, postdoctoral fellowships, and training existing staff. This demand will also impact the skillsets the Institution will look for in backfilling positions over time.

In response to a question from Senator Fischer at the hearing, you discussed the policies that your agency has in place or is developing to regulate its use of AI.

- **What specific information can you provide as to the anticipated timeline for your agency's next steps on these policies and related measures?**

The Smithsonian has existing governance processes that apply to all new technology initiatives at the Institution. AI-specific risks, such as trustworthiness, may require review and updates to those processes. The Artificial Intelligence Community of Practice (AI CoP) discussed above will be the umbrella organizing body under which additional activities can be formalized, including creation of additional policies and/or best practices for use of AI at the Institution. In addition to the AI CoP, our IT Security staff is preparing for the adoption

of the NIST AI Risk Framework to use in the evaluation of AI solutions. The NIST resources will also be leveraged by the AI CoP in establishing a policy framework for AI at the Smithsonian.

Senator Fischer

- **Can you walk the Committee through how the Smithsonian is currently using artificial intelligence? When did that start?**

Some of our earliest experimentation with AI started from a collaboration with NVIDIA Corporation in 2017 where they provided expertise and training for AI projects. The initial work trained neural networks to perform tasks like species identification of Ferns and identifying botany specimens in our collections that had been treated with mercury for pest control. We have completed additional small AI projects in collaboration with experts inside and outside the institution and co-authored publications of our work in peer-reviewed journals. Our work in AI has leveraged postdoctoral fellowships and collaborations with researchers and experts at academic organizations.

The emergence of generative AI solutions and the large language models that are optimized for various uses are being released at a rapid rate and have dramatically increased the potential for AI to be useful in our science work. In 2023, the Smithsonian Astrophysical Observatory established the “AstroAI” initiative to “Develop Artificial Intelligence to Solve to Mysteries of the Universe.” Our current use of AI is mainly on research projects; however, we are experimenting with AI to solve a variety of challenges and hope to move some of these efforts to a pilot stage later this year.

- **Does the Smithsonian have plans to expand its use of artificial intelligence?**

Yes, the Smithsonian is actively expanding the use of artificial intelligence. We are looking at solutions to enhance our operational efficiency, improve the accessibility of our collections and exhibitions for our audiences, and use it in our research to accelerate scientific discovery.

- **Is the Smithsonian exploring ways AI Can help solve specific issues the Smithsonian is facing?**
 - **If so, what are those issues?**

With over 157 million objects and specimen in our collections, we understand the incredible task of making our vast reservoir of knowledge accessible to Americans across the nation. AI has the potential to aid in the discovery of connections between objects within our collections, program content, data sets, and other Smithsonian resources. Discoveries that humans would have difficulty making at scale. We’re also experimenting with AI to accelerate the digitization of our records, create audio/video transcripts and image descriptions of our collections, and capture iconic objects in 3D. We are also experimenting with ways AI can improve metadata for our collections to improve their reach and how AI can offer new ways to approach research problems.

- **It seems like the Smithsonian is in the early stages of exploring AI and have not adapted its widespread use. Would you agree with that?**

Yes. We have leveraged AI tools for several years to support specific projects and use cases, but given the broad potential application of AI to the work of the Smithsonian, we are still in the process of scaling AI for widespread use, particularly with the generative AI solutions that are rapidly emerging and evolving.

- **The Smithsonian’s Board of Regents recommended two sites for the new museums in the fall of 2022. Unfortunately, those sites do not fit within the parameters established by Congress in the authorizing statute. We’re now over a year past the authorizing statute’s deadline for the museum’s site selection. Can you walk the committee through the Board of Regents’ plan B for selecting workable sites?**

Our process for determining the best locations was extensive, beginning with a nearly 18-month site selection process that included surveys of any property on or near the National Mall as directed by Congress. We assessed sites based on factors such as size, flood risk, accessibility, transportation, security, cost, complexity, and time to build. There was also extensive engagement with various stakeholders, to include the public, to assess the importance of various factors, such as location. It was also essential to find equitable spaces for both new museums.

After an initial review, we narrowed the list down to four locations. Three were on undeveloped plots of land, and the other was the Smithsonian Arts and Industries Building. Utilizing undeveloped plots enables purpose-built spaces and signature architecture. After reviewing the site located on lands under the jurisdiction of the Architect of the Capitol, it was clear that this site would not be practical for a museum. This narrowed the sites to three, and after extensive consultation with supporters and the advisory boards for the museums, the “South Monument” and “Tidal Basin” sites were identified as the most optimal.

While securing these sites is complicated and will require additional action from Congress, any other sites on or near the national mall would be even more challenging. The only other sites that meet Congress’s criteria are occupied by other federal agencies. Relocating an agency would add substantially greater costs and many years of additional delay. We believe these two sites are optimal and still the best path forward to build world-class Smithsonian museums.

- **We’ve discussed the Smithsonian’s deferred maintenance backlog extensively in our oversight hearings as its one of the Smithsonian’s most serious challenges. Your team began overhauling the Smithsonian’s Facilities Conditions Assessment Program in FY 2022, with changes driven by stakeholder feedback. The new program is scheduled to be fully implemented by 2026. What progress has been made in implementing the new Facilities Conditions Assessment program so far?**

The Smithsonian has made significant progress with the implementation of the new, improved Facilities Condition Assessment (FCA) Program which remains on track to be fully implemented by 2026. Improvements to the FCA process include the use of dedicated staff, a three-year rotational schedule for on-site assessments, increased stakeholder participation to improve identification of adverse conditions, focused consideration of any available engineering studies and alternate assessments, and a thorough review of documented deficiencies to verify inclusion

in Smithsonian's facilities management system. To date, the Smithsonian has completed 69% of its planned on-site assessments using the new process and has also completed 55% of its planned, updated property valuations to ensure the accuracy of the reported Current Replacement Value (CRV).

- **What kind of benefits do you anticipate seeing from this new program?**

The improved FCAs have resulted in the identification of additional deficiencies, which has led to the development of additional maintenance and repair projects to correct those deficiencies. These projects are subject to a comprehensive prioritization process to ensure the most critical projects favorably compete for the available maintenance funds. The prioritization process incorporates a new tier designation (another part of our improved FCA program) that reflects the critical nature of the building systems, the specific equipment, and the part of the Smithsonian mission which they support. The Smithsonian has also improved coordination of its maintenance projects program with the capital plan to leverage the strengths of each program and achieve improved outcomes.

In the short term, the identification of additional deficiencies has reduced our reported Facilities Conditions Index (FCI). Coupled with our more accurate and higher CRV, this has increased our Deferred Maintenance backlog. However, this short-term growth is expected to subside with a continued commitment to the ongoing improvements to our facilities management program and a strengthened level of investment in facilities maintenance and repair projects. This was evident in FY23, as the Smithsonian-wide FCI improved from 73.6% to 77.0%, largely due to the completion of several repair and revitalization projects.

- **The pandemic officially ended in May of 2023. What is the Smithsonian's current operational status? Are employees that were remote now fully back to working on site?**

All Museums and Smithsonian facilities are open and are fully operational. Smithsonian staff continue to execute our mission of ensuring the "increase and diffusion of knowledge." Given the wide variety of work Smithsonian employees are engaged in, the Institution has adopted a hybrid work model approach that includes a mix of on-site work, telework, and limited remote work. While the Smithsonian has taken steps to promote and implement flexible work practices, including the use of telework, remote work, and alternative work schedules, use of these flexibilities is balanced against the Institution's obligation to sustain its operational and business needs. As of February 6, 2024, less than 5% of our workforce (267 personnel out of 6,224 staff) have been approved for a remote work agreement. The Smithsonian is closely monitoring telework and remote work usage across the Institution and will make appropriate changes as circumstances warrant.

- **In February 2020, the Smithsonian purchased the Capitol Gallery office building in Washington D.C. for \$254 million, allowing the Smithsonian to consolidate five office spaces around the Washington, DC area, into one singular location. Is the Capitol Gallery building fully used?**

Approximately 455,000 square feet of the complex was sold to the Smithsonian in June 2020 for use as its consolidated administrative office headquarters. 283,000 square feet were already occupied by Smithsonian offices at the time of purchase. A long-term tenant, the American Farm Bureau occupies approximately 53,000 square feet in Smithsonian owned space. The building's original owner retained 157,000 square feet of office space, 20,000 square feet of retail space, and a 465-space parking garage.

In 2023, the Smithsonian completed renovations of approximately 82,000 square feet of office space, allowing for consolidation of Smithsonian staff from several leased office spaces. Staff previously located in Crystal City, D St. SW, L'Enfant Plaza, and from leased spaces within Capital Gallery were relocated into the purchased portions of Capital Gallery and those leases were not extended. Additionally, staff from the Smithsonian Institution Building (the Castle) have been relocated into Capital Gallery in preparation for the Castle revitalization. The Smithsonian is developing plans for renovation of approximately 37,000 square feet of unassigned office space in Capital Gallery to accommodate further staff requirements, including with respect to staff for the two new museums (National Museum of the American Latino and the Smithsonian American Women's History Museum). We plan to terminate additional leases in the future as we move additional staff into Capital Gallery.

- **The National Museum of African American History and Culture has required visitors to obtain timed-entry passes since it opened in 2016, but aside from that, which Smithsonian museums and facilities require visitors to have timed-entry passes?**

At this time, free timed-entry passes are also being utilized by the National Air and Space Museum. As you know, the museum re-opened several galleries in October of 2022. As expected, there was immediate demand to visit the new space, but with much of the museum still closed for revitalization, we needed to utilize entry passes to maintain safe occupancy levels. As we open more galleries, we will proportionately increase the number of entries.

Entry passes are also being utilized at the National Zoo, though they are not timed. This system has helped us to accurately gauge the number of visitors onsite to best deploy resources. It has also streamlined our parking systems and improved traffic congestion in the surrounding neighborhoods.

Occasionally, the Smithsonian utilizes timed-entry passes for special events or short-term exhibits as needed.

- **How much longer should visitors expect to need those timed-entry passes?**

The use of timed-entry passes at the National Museum of African American History and Culture continues to make sense for that museum, which has remained quite popular since its opening in 2016. Timed-entry passes minimize outdoor wait times for entry, prevent over-crowding in exhibits, and improve the visitor experience once in the museum. Over time we have enabled more walk-up access in afternoons and in off-peak seasons to ensure anyone visiting Washington, D.C. will have an opportunity to explore the museum. We will continually evaluate the usefulness of the pass system and make changes as conditions allow.

Senator Capito

- **How could AI modernize operations at each of your agencies and how could utilizing innovative technologies such as AI attract a leading-edge workforce?**

We currently have a working group (an Artificial intelligence Community of Practice, discussed above) looking at the emerging tools that can bring efficiencies across various administrative domains such as contracting, finance, and human resources with a goal of identifying a few key pilots to test out promising opportunities. We anticipate that our embrace of AI will help attract top talent as it brings exciting opportunities to participate in programs such as the AstroAI initiative at the Smithsonian Astrophysical Observatory. AstroAI strives to bring experts in artificial intelligence together with scientists to tackle the most exciting and challenging problems in astrophysics to advance our understanding of the universe and drive forward technology that will revolutionize and accelerate scientific discovery. Given the potential for AI in our research, we expect more initiatives that will bring collaborations with other academic and corporate partners that leverage the expertise of the Smithsonian.

- **Do you anticipate any procurement challenges in acquiring new technologies?**

We do recognize these technologies will need to pass a thorough security and privacy evaluation and depending on the product, trustworthiness considerations could be a barrier to adopting some solutions. We anticipate the evaluation of AI solutions will add additional time over traditional technologies, but we do not anticipate that this will be a barrier to procuring new technologies that use AI.

- **Deputy Secretary Park, with your agencies being so public facing, how could you use AI to enhance your exhibits and improve their accessibility to reach all Americans?**

One of the five focus areas of our strategic plan is to “Ensure every home and classroom has access to the Smithsonian’s digital content.” We have already begun experimenting with AI tools to address accessibility challenges with historic audio recordings and videos from our collections. We are also exploring the use of AI to generate text descriptions for website images for accessibility needs and image tags to support search engine optimization. Given the scale of the data and collections we have, we anticipate many more opportunities to emerge where AI can help improve our reach.