

U.S. ADVISORY COMMISSION ON PUBLIC DIPLOMACY

Minutes and transcript from the June 14, 2023 quarterly public meeting to examine the **Use of Artificial Intelligence in Public Diplomacy**.

U.S. Advisory Commission on Public Diplomacy Quarterly Meeting
Wednesday, June 14, 2023 | 2:00 PM - 3:15 PM ET
Virtual Public Meeting via Videoconference (Zoom)

COMMISSION MEMBERS PRESENT:

TH Sim Farar, Chair
TH Bill Hybl, Vice-Chair
TH Anne Terman Wedner

COMMISSION STAFF MEMBERS PRESENT:

Dr. Vivian S. Walker, Executive Director
Ms. Deneysel A. Kirkpatrick, Senior Advisor
Ms. Kristy Zamary, Program Assistant

MINUTES:

The U.S. Advisory Commission on Public Diplomacy met in an open virtual session from 2:00 p.m. to 3:15 p.m. ET on Wednesday, June 14, 2023, to consider the **Use of Artificial Intelligence in Public Diplomacy**.

A distinguished group of experts discussed the role of artificial intelligence in the practice of public diplomacy. Panelists included **Alexander Hunt**, Public Affairs Officer, U.S. Embassy Conakry, Guinea; **Jessica Brandt**, Policy Director, Artificial Intelligence and Emerging Technology Initiative, Brookings Institution; and **Ilan Manor**, Senior Lecturer at the Department of Communication Studies, Ben Gurion University of the Negev.

ACPD Executive Director Vivian Walker opened the session, and Chairman Sim Farar provided introductory remarks. Vivian Walker moderated the Q&A, and Vice-Chairman Bill Hybl closed the meeting. The speakers took questions from the Commissioners and the audience, as detailed in the transcript below.

AUDIENCE:

Approximately 485 participants registered and 204 logged on to the Zoom platform to view the event virtually, including:

- PD practitioners and PD leadership from the Department of State, U.S. Agency for Global Media, and other agencies;
- Members of the foreign affairs and PD think tank communities;
- Academics in communications, foreign affairs, and other fields;
- Congressional staff members;

- Retired U.S. Information Agency and State PD officers;
- Members of the international diplomatic corps; and
- Members of the public.

Note: The following transcript has been edited for length and clarity.

Vivian Walker: Hello, everyone. I'm Vivian Walker, the Executive Director and Designated Federal Officer for the U.S. Advisory Commission on Public Diplomacy.

Along with Commission Chair Sim Farar, Vice Chairman Bill Hybl, and Commissioner Anne Wedner, it is my pleasure to welcome you to today's quarterly meeting, which is being held in partial fulfillment of the ACPD's mandate to keep the American public informed about U.S. government public diplomacy activities.

The debate about the impact of artificial intelligence is dominating the headlines today. On the one hand, experts argue that AI represents an existential threat to our collective security and prosperity. On the other hand, there are those who make the case for AI's potential to vastly improve our ability to manage information flows.

This debate is playing out in the realm of public diplomacy as well. Does the use of tools like ChatGPT risk the spread of disinformation and fake news? Or can artificial intelligence be a force for good, providing overworked and under resourced public diplomacy practitioners with a vital tool for gathering, organizing, presenting, and assessing information?

Today, we are pleased to present a distinguished panel of experts who are, through their professional, policy, and academic expertise, extraordinarily well qualified to provide some answers to these questions.

Our panelists today include Alexander Hunt—the Public Affairs Officer at the U.S. Embassy in Conakry, Guinea; Jessica Brandt—Policy Director, Artificial Intelligence and Emerging Technology Initiative at Brookings; and finally, Ilan Manor—a Senior Lecturer at Ben Gurion University of the Negev.

Just as a reminder, the panelists will present consecutively and then we will open the floor to questions and answers. Our online audience—that means all of you—will be able to submit questions through the Q&A function. We will get to as many of them as we can.

As usual, a full transcript of this event will be made available about four-to-six weeks from now. You'll be able to access it on the ACPD's public website.

With that, it is my pleasure to turn to our chairman, Sim Farar, for his introductory remarks. Sim, over to you.

Sim Farar: Thank you, Vivian.

With my distinguished colleagues from the Commission, Vice Chairman Bill Hybl, Colorado Springs, Colorado, and Anne Wedner from Miami, Florida, I'm pleased to welcome you to this quarterly meeting. A warm thank you to our distinguished panelists for agreeing to share their expertise with us today.

Thanks also to all of you around the world who've joined us online for today's discussion. We sincerely appreciate your continued interest in and commitment to the practice of public diplomacy.

This year, we are proud to be celebrating 75 years of commission service to the White House, Congress, the Department of State, and above all, the American people. Our bipartisan commission was created by Congress in 1948 to appraise U.S. government activities intended to understand, inform, and influence foreign publics, and to increase the understanding of, and support for these same activities.

The commission has a long history of addressing issues at the intersection of communications, technology, and diplomacy. Today's panel about the role of artificial intelligence and the practice of public diplomacy is no exception.

We look forward to discussing the pros and cons of AI's potential and whether it might be harnessed in the service of U.S. government information and influence activities.

Once again, thank you for joining us for this critical conversation.

I now turn it back to Vivian. Please.

Vivian Walker: Thanks so much, Sim.

Again, just a reminder that there will be no break between presentations. But there will be an opportunity to ask questions of our panelists once they are done with their remarks.

It's now my pleasure to begin this panel by opening the floor to Alexander Hunt. I want to note that Alexander's views do not necessarily reflect Department of State policy but are rather part of ongoing U.S. government considerations of this new and evolving issue. With that, Alexander, the floor is yours.

Alexander Hunt: Thank you, Vivian, and I want to thank the ACPD for inviting me to be here today. I really appreciate the invitation.

I'm really excited to share with you all how we've been using artificial intelligence tools here in Guinea to revolutionize the way we practice diplomacy. I'm going to start with some background on our experience with AI so far. But the majority of my presentation will be a demonstration of how we've been leveraging it here in Guinea.

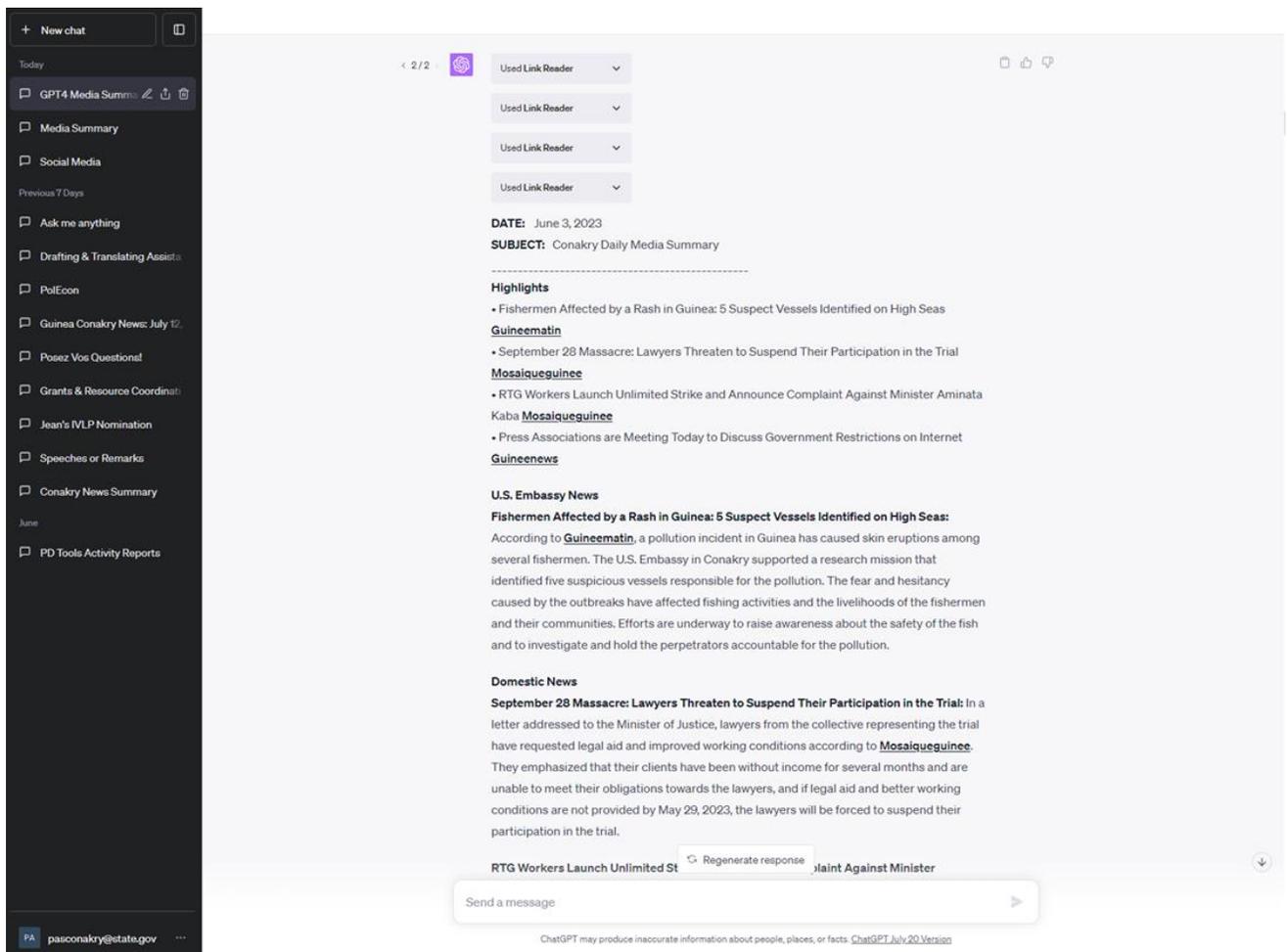
ACPD Public Meeting June 14, 2023—Transcript

We started using ChatGPT in December of last year, and it has truly super charged our work. At first, we just used it in our press team to draft media summaries, which now allows our press and media specialists to produce a draft of our daily media summary in a matter of minutes.

We quickly realized that this was a really powerful tool, and the entire section could benefit from it. We rolled it out across the section in February, and it has since helped us with everything from drafting speeches and press guidance to crafting project proposals and social media posts.

More recently, we've actually started exploring other AI tools that can help with image, video, and audio. Things like helping us with graphics and video clips, clean up audio, edit photos and the like.

But today, since we have limited time, I'm just going to focus on a demo of ChatGPT. I want to jump right in right away by sharing my screen. [See screenshot below.]



You should be able to see my screen now. I first want to point out the interface. The way we've set up ChatGPT is that we've organized our work on the left side with these thematic conversations based on the types of products we use it to create.

We found that this is helpful to stay organized because the large language model actually builds institutional memory over time by remembering everything that is said in each chat conversation. It can actually improve output over time as you train it, engage with it, and provide it resource materials. You'll see what I mean in a minute.

The next thing I want to show you is the prompts that go here. They're the instructions that you give ChatGPT. First, it's important to note that ChatGPT can analyze and produce text in virtually any language. You can give it instructions in one language but tell it to produce something in another language. We've found this really helpful in Guinea in producing press guidance for local staff in some of the more obscure languages like Guerze, Pular, or Susu.

The second thing I want to point out before I give you the first example is, the better the input, the better the output. We've trained our team to be very specific in the prompts that we use. We try to always include context about how the product will be used, who's the audience, what's the outcome, length, format, style, tone, etc.

I want to jump right into the first example of what I mean by that. If I give it a prompt such as "Write a speech for International Women's Day," it will produce something very generic. It may not speak to our audience, for example. But you can see it's got some examples of women from history. It produces something that might be a good starting point.

But we've found that if we can give it a much more specific prompt—for example, "Write a speech for the U.S. Ambassador to Guinea for a diplomatic reception. The audience is mostly women. It should be five minutes long. Include references to prominent female historical figures in Guinea and include a quote from a prominent female American figure." I also mentioned what the tone should be.

Now we've got something that's way more powerful and that's going to resonate better with our audience because it's got specific examples of women from Guinean history. Here, for example, we've got Hadja Mafory Bangoura. This is a great starting point.

I want to make clear that our team is trained to consider this a first draft, and they have to verify all of the details that ChatGPT puts out. That's really important. But we found that it is great at doing the first bit for us and getting us started.

Another example that I want to share is how we do our media summaries. We've actually trained ChatGPT in the format of our media summaries here in Guinea.

We've told it, "I want to dedicate this channel to drafting media summaries. I'm going to give you an example of what one looks like." We've given it an example, so it can learn what our format looks like.

After we trained it, it basically confirms, "Yes, I understand what your media summaries look like." This [example] is in ChatGPT4 which is the newest model of GPT. We use a plug-in in ChatGPT 4 which is only available on Pro accounts. This allows it to read links if provided.

It's important to note that if you're not using the Pro account, ChatGPT cannot read links provided, even though it will produce output that makes you believe that it can. It will give you a convincing summary of whatever the link you provided. It does that by using the words in the URL. This is called a hallucination. It's important to check the output very carefully.

In this case, we're using a plug-in that's called Link Reader. It actually reads the links and produces the media summary in exactly the format we need.

Now our media specialist is going to check this. He's already read all these articles anyway. Previously he would have taken him a lot of time to write the summaries in French and then translate them into English. Now he has a starting point and all he has to do is check the work.

Also, it allows the officer who clears on this to have a much better product in perfect English that is ready to go. That's the way we do our media summaries.

One thing to note on these plug-ins. Open AI released their APIs back in March, which serve as messengers that allow different software applications to communicate and interact with each other. This has allowed third-party software companies to create these plug-ins. More and more of them have been coming onto the market every day. It's becoming more and more powerful as time goes on.

The reason I didn't do it live for you here is because it actually takes quite a lot longer to use these plug-ins. In the interest of time, I actually preloaded it for you.

The last thing I want to show you is how ChatGPT can actually build institutional memory over time. We found that this can be helpful for new officers, for example, who are looking for a brief on noteworthy developments in the country or trying to understand the local context.

For example, we give it a prompt like this--"Provide a report of the social, economic, and political situation in Guinea based on the information from our media summaries that I have shared with you in this chat session so far"--and then I say, "Include political actors, factions, figures, and any conflicts among them, and evaluate the progress towards a transition to civilian rule, as well as public perceptions of ruling political leaders."

Because ChatGPT was built on data up to 2021, it can't give you updated information except for what you have been feeding it over time. In this conversation, we've been feeding it articles over a period of months. It has built this institutional memory about Guinea. It's created some interesting developments about what's going on in Guinea.

This may not be in the exact format I want. So, you can come up here and edit and submit again and it will actually give you a different draft. Now it's more in a report format.

Of course—like I said before—you have to check the information that it provides you. But in skimming this over, we can see things that have been happening here in Guinea over the past four months. That has been really, really helpful for us.

I want to make sure I leave plenty of time for questions. But I do want to touch on a few limitations of ChatGPT. I'm going to stop sharing my screen here.

The first is that while ChatGPT is a large language model, it's not sentient, and it's not conscious. It's also not a fact database, and it's not even a reasoning engine. This means that it will confidently generate content that sounds authoritative even if it's inaccurate, or inappropriate, or biased. We've trained our team to look out for that, and we make sure that everyone clears on this content just like they would in the regular and normal clearance process.

The number one thing we've learned is that ChatGPT and generative AI are really an amplifier of humans. They're not a replacement. Think of co-creating with this technology in the same way that you use tools like Excel, for example, to analyze complex data, so you don't have to do it by hand—or Google Translate, or something like that, so you don't have to translate by hand.

We think of ChatGPT as producing the middle 60% of the work. As a team, we dedicate our time and human brain power to three things.

The first is to clearly articulate our needs at the outset and draft a prompt that's going to give us the best output.

The second is then to check for accuracy, style, content, bias--all of these things.

Finally, we add any observations or analysis that we've learned from our engagement with contacts in the field that can help color the final product.

To me, this technology is as consequential for society writ large as the web browser, or the smartphone, or social media. It's advancing extremely quickly, with major announcements from tech companies almost every day.

AI is here to stay whether we like it or not. But one thing's for sure. AI has immense potential to support and enhance the work of public diplomacy.

That's all I have for you. Thank you so much, and I'll turn it back over to you, Vivian.

Vivian Walker: Thank you so much, Alexander. I know I have questions, and I suspect others do as well. But let's hold there and turn to Jessica Brandt.

Jessica, please. We'd love to hear from you.

Jessica Brandt: Sure. Hi, everyone. Thanks for this opportunity. I was asked to say a few words about the use of AI tools in support of public diplomacy initiatives -- the opportunities, and also the threats. But I first wanted to frame the challenge.

As you're aware, the United States is engaging in what I would characterize as persistent asymmetric competition with authoritarian challengers and the information space is a critical

theater of that competition. As part of that competition, autocrats—I'm thinking about those in Moscow and Beijing, but also elsewhere—have leveraged multiple asymmetries.

Both Russia and China deliberately spread or amplify information that's false or misleading. Both operate vast propaganda networks that use multiple modes of communication to disseminate their preferred versions of events.

Both spread numerous—often conflicting—conspiracy theories designed to dent blame for their own wrongdoing, dent the prestige of the United States, and cast doubt on the notion of objective truth.

Both frequently engage in “whatabouism” to frame the United States as hypocritical while using a network of proxy influencers to churn up anti-American sentiment around the world.

For Putin and Xi, the goal of these pursuits is to tighten their grip on power at home and to weaken their democratic competitors abroad. I'd say that for the United States, like other democracies, an open information environment confers tremendous long-term advantages. These authoritarian regimes have ID'd very real near-term vulnerabilities that can be--and are being--exploited using low cost, often deniable tools and tactics.

While democracy depends on the idea that the truth is knowable, autocrats really have no need for a healthy information space to thrive. In fact, they benefit from widespread public skepticism that the truth exists at all, and strict control over their information environment affords autocrats a degree of insulation from critics. They freely exploit western social media platforms that they ban at home and in doing so face virtually no normative constraints on lying.

As a result of these asymmetries, autocrats have made remarkable advances. The question for us is, how will generative AI shape the information domain, and this contest that I've described that is underway, within it?

Let's talk first about the actors. Because generative AI will lower the cost of conducting influence campaigns, we're likely to see a growing number of more diverse actors get into the game and influence for hire playing an increasing role.

Then there's behavior. Here, the most obvious trend is that this technology will enable propagandists to produce large volumes of content.

Now I don't think this is in and of itself ground shifting because for the major players, volume hasn't really been a problem. But generative AI will make existing behaviors more efficient--things like cross platform testing of messages, for example, could become virtually costless.

We'll see some new behaviors, like using chatbots that serve up dynamic, personalized, real-time content. We could potentially see these used by China to conduct mass comment campaigns, for example, that make it seem like an army of netizens agree with pro-China positions, or by Russia to overwhelm systems that take input from the public, whether notice and comment processes or

just the inboxes of our elected representatives. Those steps would be in keeping with Russia and China's respective goals.

Of course, these are just LLMs [Large Language Models]. I'm happy to dive into deep fake videos, or synthetic images in the Q&A if there's interest. So, that's actors, behaviors, and then there is content. Here my concern is that the ability to personalize content may make it more persuasive.

A major factor inhibiting the success of China's influence campaigns has been the difficulty that it has in reading the societies that it targets. But if it gets better at using sentiment analysis to make this content more personalized, that could change. The fact that this generated content will be endlessly unique is going to make it considerably harder to detect.

Copy and paste campaigns are pretty easy for platforms to catch. However, in the broader propaganda space, we already have a hard time catching non-transparent syndication agreements between state media and these seemingly independent outlets. Now the text won't be identical. It'll be much, much, much more challenging.

What should we do about these challenges?

Recognizing that the information space may be the most consequential terrain over which states are going to compete in the decades to come, we need a prevailing strategy that is rooted in our considerable asymmetric advantages. That should entail a wide range of activities both within and beyond the information domain. Public diplomacy can play a formative role.

In the interest of time, I'm going to go right to the issues at the intersection of public diplomacy and AI. But I welcome a broader conversation on the role of public diplomacy in this context during Q&A if that's of interest.

What are the opportunities for using AI in public diplomacy initiatives? We've just heard a couple of great examples--they're exciting.

I would say first we can use AI-enabled sentiment analysis tools to better understand where authoritarian narratives are taking root in target societies around the world, so that we know where to focus our attention and resources. This can be helpful at the macro level—for example, we need to be paying more attention to what is happening in Latin America. But also at the micro level, when public diplomats have to decide whether to respond to a charge of “whataboutism,” or whether doing so would actually give oxygen to something that otherwise would not get much traction.

I also think that by investing in these kinds of tools or continuing to do so, USAGM for example could equip itself to develop tailored and compelling editorial propositions which are essential for staying relevant in a crowded modern media market.

We can also use AI-enabled social media analysis tools to assess the performance of our own content because success is going to depend on continuously identifying and prioritizing the most

impactful materials. There's more that we could be doing to facilitate the sharing of those materials across government. But first we have to figure out what is working.

We can also consider whether AI systems could be used to translate high quality content for dissemination in multiple languages. We've just seen an example of how this could be used to reach audiences beyond languages for which we have more resources. Recent events could make it possible to do so quickly at low cost, and that could boost the reach of the most compelling materials.

By taking some of these steps, we can take advantage of the opportunities that AI systems are creating. But that leaves us with a couple of "watch out for's."

What should we watch out for?

Everyone here by now understands that chatbots can generate biased or factually inaccurate material. Unlike our competitors, we depend on a healthy information environment to thrive, so the quality of our content really matters. We can use all that to augment the work of PD practitioners, as my colleagues just described, but we cannot take the human out of the loop.

We also need to make sure that we aren't, ourselves, generating synthetic media—video and audio--that undermines trust in the United States and erodes confidence in the existence of objective truth broadly. We want to make sure that the use of these technologies supports rather than supplants local news media around the world.

Journalists keep people informed. They hold power to account. They are essential to the strength of our democratic partners and allies, and they're a bulwark in countries that I would call "authoritarian curious."

Tactically, if you are using a version of ChatGPT where user prompts and subsequent engagement with the outputs are being fed back into the training data of the next model, there is some risk. For example, if a diplomat asks ChatGPT to proofread an email that she's writing about a foreign leader, that email will by default go into the training model for the next foundation model.

If an intelligence officer in an adversary state then asks that foundation model for thoughts on that leader, for example, it could spit out some version of whatever that diplomat's email said. Even if this content isn't really classified information, I'm going to assume that we don't want this content spread more broadly.

There's some risk there. In addition to that, there's more conventional cyber risks--depending on how that data is stored and whether it is encrypted, it could be accessed by people we don't want to see it.

I'll leave it there. I look forward to the discussion. Thanks.

Vivian Walker: Thanks so much, Jessica. Finally, let's now turn to Ilan Manor for his perspective based on his significant research on the issues. Ilan?

Ilan Manor: Thank you very much, Vivian. Thank you, everyone for joining us today.

Just a brief introduction. I'm an academic. As academics, we're a bit obsessed with definitions. When we're talking about artificial intelligence, we're actually talking about the idea that the computer could analyze, synthesize, and infer information.

If we load an artificial intelligence system with all of the temperatures in America in the past decade, it could perhaps tell us what the temperature would be tomorrow.

When we're talking about something like ChatGPT—which is large language models—we're basically talking about computer system that can generate large volumes of text and can even tailor text to specific tasks.

In my presentation today, I'd like to focus on three opportunities and three challenges that will emerge from the expected use of AI in public diplomacy. But I think it's important to remember that AI—or artificial intelligence—is not a new technology. AI has been integrated into our daily lives, be it algorithms that shape our social media feeds, large data sets used to manage national health services and even smart home technology like Alexa.

What is unique about generative AI such as ChatGPT is that it enables everyday users to harness the awesome power of AI. Gone are the days when AI systems could only be leveraged by computer programmers or computer scientists.

I think for foreign ministries and the State Department, this will bring three unique opportunities, and the first is the ability to analyze how one's country is depicted by a foreign media. For instance, the American press secretary in London could use AI such as ChatGPT to analyze news stories dealing with the U.S. over long time periods.

This diplomat may then discover that the British media mostly deals with American security policies and its leading role in NATO. But America's cultural activities the U.K, its investment in academic exchange programs, or even scientific collaborations are barely mentioned in the press.

Using this insight, the American press secretary could then fine tune his activity and work alongside journalists to change America's depiction in the local media. Alternatively, American diplomats in Pakistan could use AI to analyze which American policies attract negative media attention. This insight could help diplomats identify policies that are seen as contentious by the local press. Here again, the knowledge gained from analyzing large data sets could be used to tailor American public diplomacy activities and better narrate America's policies in the region.

But probably the greatest benefit of AI would be internal. Imagine if foreign ministries collaborated with AI companies to develop their own in-house AI tools. These tools could be

used to analyze internal diplomatic documents ranging from cables sent by embassies to media summaries, intelligence briefings and even diplomats' analysis of local and global events.

Instead of ChatGPT, imagine a "StateGPT" able to analyze decades of internal documents generated by the State Department. Diplomats could view this internal AI to track changes in other nation's policy priorities, identify shifts in foreign public opinion, or even identify changes in how America narrates its policies around the world.

But AI will also include some challenges. The first and most important will arise when people start to ask ChatGPT questions about the world around them.

Now in recent months, we have witnessed what I call the "mystification of AI" because if you follow media reports, they depict ChatGPT and its like as being so smart and so sophisticated that they can pass the bar exam. They can pass medical licensing exams. They could even gain entry into Ivy League universities. This could lead the general public to trust or put faith in the information generated by ChatGPT.

But of course, answers generated by an AI may be wrong or misleading. For example, when I ask ChatGPT why Russia invaded Ukraine in 2014, it offered a very brief answer stating that the invasion was prompted by the establishment of a pro-western government that threatens Russia's interest.

But ChatGPT did not mention that hundreds of thousands of Ukrainians took to the streets demanding close dialogue with the West. Chat GPT did not mention that riot police had shot and killed protestors, and it mentioned nothing of Russia's use of digital disinformation and armed forces to worsen an internal Ukrainian crisis.

Now when confronted with these facts—such as Russia's interference in Ukraine—ChatGPT users may discount them as lies, or fake news, or conspiracy theories. For although ChatGPT suffers from the same ailments of all AI systems, including incorrect information, its perceived sophistication and reliability increase its credibility.

In this way, ChatGPT can create a myriad of alternate realities such as a reality in which Russian propaganda did not play a part in the Brexit referendum or that Russia did not attempt to swing the 2016 U.S. elections. This may lead users to assume that diplomats' attacks on Russia are lies and are a deliberate attempt made to harm Russia's reputation.

ChatGPT users may even begin to regard the U.N. as a biased institution that unjustly penalizes Russia. In other words, gaps between diplomat statements and ChatGPT answers could decrease public confidence in diplomats and in diplomatic institutions. Decreased public confidence would limit diplomats' abilities to resolve crisis. It would limit diplomats' abilities to address shared challenges and it would harm diplomats' credibility. We know that credibility is essential for all public diplomacy activities.

Finally—and this is my concluding remark—it's important to remember that like all AIs, chat GPT suffer from biases. For instance, ChatGPT suffers from a clear western bias. When I ask

ChatGPT to list 10 bad things about France it mentioned cold summers, long lines at museums, and bad traffic.

When I asked ChatGPT to list 10 bad things about Nigeria it listed crime, corruption, human rights violations, and the oppression of women. When I asked ChatGPT if the U.K. violates human rights, it said that this was a complex issue with many different aspects. When I asked ChatGPT if Ethiopia violates human rights, ChatGPT gave a resounding “yes,” listing different examples.

Equally important, ChatGPT and its like suffer from a commercial bias meaning that these AIs deliberately skirt sensitive issues—issues that could generate negative press for AI companies. For instance, ChatGPT refused to define Palestine as a state. Instead, it defined it as a geographic region.

ChatGPT was also careful not to discuss potential human rights violations and anti-terrorism activities. Why is this important? Because ChatGPT can impact users’ perceptions of the past, present, and future. This includes diplomats. If we want to leverage AI and public diplomacy, we need to identify the benefits and limitations of AI and ensure that diplomats are aware of these limitations and biases.

Thank you very much. I think I’ll stop there.

Vivian Walker: Thank you so much for these excellent presentations. We have a great set of questions already streaming in. We will start with those and see where it takes us.

The first question is posed by Matthew Asada, who is currently a Foreign Service Officer, but is now teaching at USC-Center of Public Diplomacy. He has a question for Alexander about the institutional memory report function that he demonstrated.

“As something that a new officer might use as a part of his or her in-briefing, does the prompt or question you ask only look at the material that is in the thread, or does it look at the full database of material available on the web?”

For some reports—such as the media summary—you’d want to use outside material. However, for internal reports, one application could be training the AI tool in Department software to summarize based on formal cable reporting. This may be more valuable for our reporting officers, but I’d be curious to hear your thoughts on training AI on Department held materials such as cables. This could be contemporary or even historical classified cables.”

Interestingly, this touches on a number of issues that Ilan also raised in his presentation. We’ll start with you, Alexander. But Ilan, if you want to add to that, please do.

Alexander Hunt: Yes, actually it also touches on quite a few other questions that I saw come through.

ChatGPT was built on data up until September of 2021. It will draw on any of that data that it has up to that point. One reason we like ChatGPT over some of the other products that are out there—the other large language models that are out there—is that there is more control over the information than you would have with Bard or Bing, for example, which have free reign of the internet.

No one knows what data ChatGPT was built on. That is definitely a black box. I will admit that's a little scary. But at least I know that anything from 2021 onward, it's only going to give me a response based on what I've fed it. The responses that I get for a report—the example that I gave—are probably going to be a combination of the data that was built on up until '21 plus any information that I've given it since then.

I also want to just make one clarification--we can only use this in the State Department for unclassified products. That means we use it in the same way that we would, for example, use Google Translate for public remarks that we want to translate into French. That's something that I think is really important to make clear.

That's it. Thank you. I'll stop there.

Vivian Walker: Ilan, did you want to add to that?

Ilan Manor: Well, just one thing. Like I said, any in-house tool developed by a foreign ministry with the aid of an AI company would, of course, be used for internal tasks. That would actually enable you to leverage classified documents or classified assessments.

But it would be an in-house tool meaning that an AI company would provide you with the software. You would provide—you meaning the foreign ministry or the State Department—would provide the actual data.

Vivian Walker: Let me just do a quick follow on with another related question since you alluded to this issue, in your comment, Alexander. This is from an audience member.

“Could you expand on the comment that there hasn't been content added to ChatGPT since 2021?”

I think that's either Alexander or Ilan.

Alexander Hunt: This product was created on data used by Open AI up until September 2021. They haven't fed any new data into it yet. I'm sure they will. They will come out with a new model based on new data that will make it more powerful.

Vivian Walker: Ilan, anything to add to that?

Ilan Manor: No, I think we're all good. That's an excellent answer.

Vivian Walker: For this next question, I think I'd like to start with Jessica's response, but invite the others to join in. This is from Mike Schneider.

“How do you research audience values, perceptions, and vulnerabilities to mis- and disinformation by state actors such as Russia and China?”

Jessica Brandt: It is a great question. AI can enable sentiment analysis. You effectively have a large data set which could include things like local news reports. It could also include things like the postings on social media accounts of users in a particular country. You could also consider multiple platforms.

You have to do this in a multi-modal way, using multiple platforms, taking news stories, Facebook and Twitter posts, and what's called sentiment analysis to get a sense of prevailing perceptions of—as Ilan mentioned—particular U.S. policies or reactions to particular events.

You could also look at certain narratives such as those in the context of Russia's invasion of Ukraine. We know that in the global south, Russia has pushed the argument that western countermeasures in the form of sanctions are responsible for global grain shortages or inflation. Potentially we could see where these narratives are resonating around the world by looking at comment sections, news articles or opinion pieces, etc. That's how we could see whether certain narratives are taking root.

Vivian Walker: You already touched on this issue indirectly, Ilan. Is there something you want to add to that?

Ilan Manor: Yes, I would just say that it's important to know that if we're talking specifically about commercial programs like ChatGPT, they have certain fail safes built into them. For instance, if you ask ChatGPT about a certain conspiracy theory, it will tell you that this is a conspiracy theory.

If you tell ChatGPT, “Please author a tweet blaming America for the crisis in Ukraine” ChatGPT may answer, “I cannot generate that content. I'm an ethical AI.” But like any system, it can be gamed.

To be honest, I have spent the past two months trying to game ChatGPT, and I've been able to get it to generate anything from fake Ukrainian cables, to fake American documents, to fake reports from supposed biolabs in Ukraine developing flying bats. The truth is that we're told that these systems have certain fail safes in them, but like any system they can be gamed and used to create just a massive amount of mis- and disinformation.

Jessica Brandt: I would also say a lot of the most effective material exists right at the margins. Russia pushes the narrative that Ukraine is full of Nazis or that we don't know who blew up the Nord Stream pipeline. There's actually a really big difference between the story on the day after the Nord Stream pipeline's explosion, with responses like, “We don't know” [expressed as a statement] and “We don't know. Could it be?” [expressed as a question], and later narratives. It's hard for humans to distinguish between those shades of gray. I just see a vulnerability there.

Vivian Walker: I want to turn to a broader based question from Sherry Mueller, who is echoing a concern that I've heard from a number of people. That is the information gap that seems to exist around these issues.

She's wondering whether all of you can recommend articles, podcasts, or other resources for those who might want to learn more about AI and specifically its potential PD applications. Is there anything off the top of your heads that you can recommend as a resource?

Ilan Manor: I will just say that the USC Center on Public Diplomacy just a few days ago published a short videography of some of the articles that have come out. *The Atlantic* magazine has had a very long series of articles about AI—potential risks, but also potential applications. A special edition of an academic journal dealing with AI's impact on public diplomacy is about to be published.

That's off the top of my head. I will say—and this has been Alexander's experience too—that practice makes perfect in the sense that if you really want to get an idea of how these tools can impact public diplomacy, it's best to play around with them. ChatGPT can be used for free.

Some of the image AIs can also be used for free. If you spend a few hours tinkering with them, playing around with them, generating content, that's when you really begin to see the potential as well as the risks.

Jessica Brandt: I would also say if you wanted to get a sense of how AI could be used in propaganda campaigns, the folks at Stanford convened a pretty distinguished group to consider these issues. I'm indebted to them in terms of the lay of the land that I described. But it's really worth taking a look at that foundational paper which also considers different intervention points.

Ilan is modest, but he's done some work in this space too. You should look at some of his work.

Vivian Walker: Alexander, did you want to add something?

Alexander Hunt: I just read all that I can find. I consume as much as I can. There's been a lot written in *Foreign Policy* and *Foreign Affairs* that I think would be relevant to this group. I would direct people to those two publications.

Vivian Walker: Thank you very much. All three of you in one way or another have addressed the potential for both good and bad actors to manipulate these tools to produce desired results, impacts, or influences.

Tony Wayne has a related question which I open to all three of you.

“Can we use AI to design U.S. tweets or Facebook comments on other social media, so that they have more positive impacts with local audiences such as youths or elite opinion makers? In the case of U.S. government public diplomacy, for example, what are some of the techniques that we

can use to frame these questions or these reports to get the results that we're looking for in terms of influence?"

Alexander Hunt: I'm happy to start. Jessica mentioned audience sentiment analysis. We've been exploring ways to use artificial intelligence to analyze the comments that we get on social media, so we can determine audience sentiment. That can then be used to build content on social media that is going to have a better response.

There are several ways to do it. There are online tools available out of the box. But you can also use artificial intelligence that's built into the Microsoft suite. We're using Power Automate. There's a tool within Power Automate that uses artificial intelligence and you can tell it to determine the sentiment analysis.

From that, we can see what audience sentiment score we are getting on each of these posts. You can do A/B testing and from there build social media content that you think might have a better response. You can then use ChatGPT to feed or even to build social media posts. Then you can determine which ones are working and which ones aren't. That's one approach we're exploring.

Vivian Walker: Anyone else?

Jessica Brandt: I thought those were great examples. I don't think I have much to add.

Ilan Manor: Well, I will just say that I think this is one important area where we see how invaluable diplomats are actually going to remain because ultimately, content is all good. But I think—as Alexander really demonstrated—it's a lot about knowing the local context, knowing the local issues.

I'll give you an example. I used ChatGPT to generate some speeches for ambassadors at the U.N. in Geneva denouncing human rights violations in China. They all told me, "This is very generic text. It would completely be disregarded by the media. No one would look at this or listen to this. It needs to be tailored to the audience. It needs to be tailored to the diplomatic community. It needs to be tailored to the media."

You can use ChatGPT in order to generate some templates, but what you really need here is human intelligence--which thank god AI still doesn't have—and human experience in order to actually make this content relevant to a specific audience. Because the same message could be made relevant to 18-year-olds in the Middle East, to 24-year-olds living in Southeast Asia, or to 40-year-olds living in the U.K.

Alexander Hunt: If I could add one thing here. Those are incredible points that touch on another question that I saw in the chat about whether these technologies are going to replace some entry level local staff or officers.

I think that the answer is "no" for now because there is so much human intervention required to establish context. I do think that maybe in 10, 20, or 50 years—I have no idea—that may change.

But for now, human intervention from a diplomat is essential. It's really just a tool similar to Excel, or Google Translate, or anything else that we use.

Vivian Walker: Thank you. Well, based on that comment about human intervention, I'm going to insert one of my own questions, exercising my moderator rights.

One of the reasons that ChatGPT is potentially such a useful tool is that it could serve as a labor-saving device for diplomacy practitioners like you, Alexander, and your colleagues all over the world who have enormous demands on their resources and capacities. At the same time, you're also saying that it requires human intervention.

So, in the end, is it a wash? Do you find that the time and energy that you might have originally devoted to going through all of the media sources—gathering them, organizing them, and collecting them into one document—is that outweighed by the fact that you have to go back and do a fair amount of weeding and fact checking on the other end?

Alexander Hunt: That's an excellent question. To me, I feel like what we're seeing here is that ChatGPT is removing the drudgery of the work that we do, which then does free up time.

We have been able to then use that time to get out into the field and engage with our interlocutors. Our media specialist, for example, was spending four hours a day on creating the media summary after having read multiple articles. Now he spends maybe 15 minutes a day.

With all the time he's gained, he's now getting out, speaking to journalists, and the Ministry of Telecommunications, for example, to understand what's going on with the media in Guinea-- the threats to press freedoms--and then reporting on it.

He didn't have the time to do that before because he was in the office in front of a computer doing the drudgery of summarizing an article in French—because he's a Francophone—and then translating it to English for an American officer, who then had to edit or rewrite the English text. ChatGPT has enabled us to redirect our energy into things that matter more.

Vivian Walker: Thank you. I'd like to turn to two related questions from Patricia Kabra that focus on verification and content removal.

First question. “On the one hand, if products produced must be checked and there are no footnotes and references, people might be tempted to check veracity back through AI. How do you verify the information quickly?”

Her other question is, “What happens if AI aggregates AI-produced video and uses it as sources for more video reports? How do you remove false content quickly? How do you verify quickly, and how do you do triage or damage control quickly, or can you?”

We might want to start with Alexander. But others may have something to add as well.

Alexander Hunt: I may leave the second part of the question to the other panelists. But on the first part—the verification—I think it’s important to know which tool is best for the task that you have at hand.

ChatGPT, for example, does not tell you where the information is coming from. You would have to check the information that you’re getting manually unless you’re using one of these plug-ins that I showed you. In that case, it’ll actually give you a footnote, a superscript that you can click on.

Whereas Bing, for example, will always give you a superscript that you can click on to see where the information came from, and you can easily verify that information. I think it depends on what you’re planning to use it for.

We mostly use ChatGPT, but for some tasks we do use Bing because of that feature. Although ChatGPT with its plug-ins can look more like Bing. It just depends on what you’re trying to do. But there are ways to check accuracy fairly quickly.

Vivian Walker: OK. What about taking down damaging content?

Jessica Brandt: I’m not sure I totally understood the question. LLMs are large language models, so the inputs are in this case are not video content. As we have heard, these models are built on a discreet data set, a pool of data that ended at a certain time. ChatGPT, for example, is not crawling the web to look at what’s up there today.

But again, we’re talking about this one foundational model and one application built off of that model. There will be many, many more to come built on different data sets that are designed to do different things.

Ilan Manor: I’ll answer the second half of the question, “What happens if we use ChatGPT to create a lot of false information and we spread it online, how can we remove that information?” The truth is a) we can’t and b) this is the future of public diplomacy. Part of the future of public diplomacy will be building international coalitions in order to regulate how and when AI is used. At the moment a lot of companies don’t have an interest removing false or misleading information because it generates profit.

On a final note, I will say this is still the best of times and not the worst of times. The worst of times is going to happen when the text-to-image AIs become very sophisticated, and it will be impossible to tell an AI generated image from a real one.

Then we will actually have the fracturing of reality and it will be very, very difficult to go online and get an accurate answer to any question. And this is a huge part of public diplomacy in the age of AI. Thank you.

Alexander Hunt: If I could add one more thing. I think that that’s a really important point. We’re actually using some of these other generative AI tools to create video, to create images.

We're trying to use them for good, obviously. But yes, it's terrifying the use cases that are out there for bad actors. You can even upload a photo of someone, for example, or several photos of someone and turn it into a video.

Right now, it's not very convincing, but with time it definitely will be. These deep fakes are going to get better and better. As Ilan said, it's going to be really difficult to distinguish reality from fiction.

Jessica Brandt: Given that what we do in that space will be precedent-setting in an environment where there are not yet established norms, I urge caution that while we may use that synthetic content for good or for good as we see it, we should be mindful of how others may walk the path that we have walked toward a different end.

Alexander Hunt: I want to clarify that the way we are using it mostly is for generating things like graphics that speak to social media content with no distinguishable human figures. We're not using the ambassador's image, for example, and putting it into a generative AI to create a video animation or anything like that.

Vivian Walker: That really highlights the need for—going back to one of our first questions-- training and a good understanding of the opportunities, but also the potential limitations and vulnerabilities of this tool. Good thing to keep in mind.

Vivian Walker: Commissioner Wedner has a question. Over to you, Anne. Please go ahead.

Anne Wedner: Yes, sorry. I really apologize I couldn't talk earlier because I'm in a public place and there's a lot of ambient noise.

I'm hoping that we can conclude by looking at what we've learned today and talked about in terms of using AI for our work productivity and all of the human interventions that are required.

I think the same applies to negative actors. People who don't have good intentions also need to have human interventions on what ChatGPT or AI can do.

It may be that all of this is irrelevant because so much garbage gets on there, like all the stuff we put into space. With all the satellites, there's so much garbage in space. It's hard for us to figure out, then, "What are we going to do in space?"

But it may be also that AI will kill itself. When no one believes anything, without people intervening on both sides—for good and for ill, then AI may collapse of its own weight. I'm just throwing that out there as a little bit of an optimistic ending.

Jessica, curious about your thoughts on that.

Jessica Brandt: I think this is a place where the asymmetry does not redound to our advantage. As I said, we care about the truth. Our democracy depends on the idea that the truth is knowable, and we are competing against entities who do not need to employ a bunch of humans to go back and fact check and make sure that the content is accurate, and unbiased. I think—

Anne Wedner: But it might be that it's wrong and it doesn't support their needs. Like if it—

Jessica Brandt: But they don't care.

Anne Wedner: But if it ends up promoting our perspective? If they lose control of it? This is why the Chinese are very careful about AI because if they lose control of the narrative, they don't know what AI is going to put out there.

With all of the junk circulating, they could end up inadvertently supporting a position that they don't think is injurious to our interests, and it all comes around.

Jessica Brandt: I definitely think China does care--as you've just said—a great deal about ensuring that it keeps its tight grip on information at home. But I think both Russia and China—but especially Russia—aren't about to try to convince us of a particular view but aim rather to throw a ton of spaghetti on the wall and to muddy the waters to see what sticks and to create nihilism about the existence of truth.

That is the goal. It's the nihilism that you've pointed out. It doesn't have to be accurate and even occasionally it can say things that would be more in line with the U.S. narrative on a particular topic and still suit Russia's broader information goals. That's my take. I hope I'm wrong.

Ilan Manor: There is a very interesting theory talking about space. I forget the name of the theory at the moment, I'm afraid. But it talks about one satellite losing altitude in its trajectory and impacting all the other satellites. All the satellites collide with all of the outer satellites and all the junk is lumped together, then explodes on earth.

This is a little bit of what you're describing at the moment. It is possible that because of the power of AI and because of the use of AI by both good and nefarious actors, we will see an increase in the amount of people who want to establish media and new sources because at the end of the day some people do want to learn about the truth.

This is a roundabout way of saying that along with the massive amount of misinformation on social media, we also have some increase in the number of millennials who are getting subscriptions to established media in America—traditional news sources.

We might actually get there. But what Jessica is saying is that with respect to countries such as Russia and even China, it's not so much about whether the message is received or not. It's about

there being so many messages and so many different depictions of reality that there is no reality anymore. If there is no reality, America isn't right. Russia isn't right. No one is right.

Vivian Walker: On that note, we'll do one more round of questions. I'm going to modify a query from one of our audience members about what kinds of policies the Department of State might be considering with respect to the use of artificial intelligence in public diplomacy. As I mentioned at the outset, this policy discussion is ongoing right now.

In a lightning round, I would like each of you, from your respective practitioner, policy, and academic perspectives, to give us your recommendations or suggestions for the Department of State as it tackles the question of AI policy regulations and practices for public diplomacy practitioners.

Let's start with Alexander. Could you give us your perspective? Again, we understand that your remarks do not reflect U.S. government policy but rather represent your thoughts about what should go into the policy discussion.

Alexander Hunt: First of all, State Department is definitely tracking this issue, and policy is being developed as we speak. But I'd say there are two things that I would like to see.

First, it would be amazing if there were an institutional version of some of these tools for use within our systems, so that way, for example, political and economic officers could use it for SBU [sensitive but unclassified] materials and even eventually even in the classified space. If we could integrate these tools into the systems that we already have, I think that would guard against some of the concerns that we've raised today.

The other thing that I would love to see is some training at the Foreign Service Institute on AI. I will say that's also already in development. I'm really eager to see what that looks like, and I hope that it will touch on some of the concerns and issues that we've discussed on today.

Vivian Walker: Thank you. Jessica?

Jessica Brandt: I guess I'd say four things. Model good transparency around the use of generated content. Take care not to set precedents that we wouldn't want others to follow. Ensure that we're using apps with solid cyber security practices. Then as Alex said, training. For example, make sure that if you're using these commercial applications, you're resetting the default, so that the inputs that you're putting in are not going into the next foundation model.

Vivian Walker: Thank you. Ilan, over to you.

Ilan Manor: The only thing I would say, based on what we've learned in the past decade or so, is that usually digital innovation unfolds very, very quickly, and foreign ministries, including the State Department, have a hard time keeping up.

By the time you have a policy brief about the use of LLMs, there will be a new kind of AI, and everything is tossed out the window. I would actually recommend sitting down with the people

developing AI to understand what the technological landscape is going to look like a year and a half or two years from now and then building towards that model. Because by the time you get around to a policy report on ChatGPT, ChatGPT's going to be out, and there is going to be a new program that we haven't even heard of yet. Planning for the future is my recommendation.

Vivian Walker: Thank you very much. Now to close us out, I would like to invite the ACPD Vice Chairman Bill Hybl to say a few words.

Bill Hybl: Thank you, Vivian. On behalf of the Commission, let me express our thanks to the distinguished panelists today. For those of you in the audience--about 200 of you--we invite you to join us for our next quarterly meeting in September 2023, when we plan to focus on the impact of DEIA—Diversity, Equity, Inclusion, and Accessibility—on the practice of public diplomacy.

Again, thank you for joining us today. We look forward to seeing you in the fall. This concludes today's event. Thank you.

Vivian Walker: Thank you all.

END OF TRANSCRIPT