s47C & s33

s47C & s33

s47C & s33

s47C & s33

s47C & s33

s47C & s33

s47C & s33

s47C & s33

s47C & s33

s22

Nov 01, 2023   View in browser

BY **DEREK ROBERTSON**

Bletchley Park, near Milton Keynes north of London. | AFP via Getty Images

The United Kingdom's much-hyped AI Safety Summit kicked off today — and it's already revealing what world leaders hope for, and fear, in an AI-powered future.

The most concrete takeaway from the first day of the two-day summit is the "Bletchley Declaration," named after the Bletchley Park estate where the summit is hosted (and where Alan Turing pioneered the codebreaking that helped the Allies win World War II.) The document, with 29 signatories (including China) uses mostly boilerplate language to address both existential concerns about the threats posed by powerful AI models in the future, and more quotidian policy worries about how AI might supercharge the more harmful or biased parts of existing bureaucracy.

But if you look at the individual statements from attendees of the summit, a clearer vision begins to emerge for where their priorities lie, and what they see when they are projected five, 10, or even 100 years ahead.

Starting with the hosts, U.K. Prime Minister Rishi Sunak's fear of an AI apocalypse is well-documented. Fueled by speculation among some academics and high-level AI developers (including Elon Musk) that a "God-like" AI could potentially wipe out the human race, Sunak favors

1

reining in super-powered frontier models, and the Bletchley Declaration nods to that (theoretical) extinction-level threat.

The United States' representative does not share that priority, or at least not to the same extent. As POLITICO's Vincent Manancourt, Eugene Daniels and Brendan Bordelon reported today, Vice President Kamala Harris is more worried about the threats AI poses to society right now.

"Let us be clear, there are additional threats that also demand our action. Threats that are currently causing harm, and which to many people also feel existential... when people around the world cannot discern fact from fiction because of a flood of AI enabled myths and disinformation. I ask, is that not existential for democracy?" Harris said.

That is to say, the harm generated by humans using AI might be worth more scrutiny at this moment than those generated by machines at some undefined point in the future. (This also happened to be the implicit message of the most recent "Mission: Impossible" film, as President Biden might have done well to read in a DFD edition from earlier this year.)

This philosophical split isn't a matter of excluding or ignoring one type of risk over the other *entirely*, as the Bletchley Declaration nods to both. But governments have a limited amount of political capital, and sway — and the regulations on which they'll decide to focus are far from certain at this point. Companies are afraid of that uncertainty, and some are bristling at how a patchwork of regulations focused on different priorities might complicate what they do.

POLITICO's Rebecca Kern kept tabs on Meta president of global affairs Nick Clegg for DFD yesterday as he made the media rounds pre-summit, and he said the race by world governments to regulate AI is confusing the industry.

"All of that doesn't quite seem to fit together," he said at an AI Fringe event, hoping this week's summit will lead to governments at the very least being "corralled in a similar direction." (He also said he hopes that governments can agree to some common definition of what a

"frontier model" actually *is*, a concept still largely up for debate even as Sunak et al. discuss how to regulate them.)

Meanwhile, developers, academics, and activists who aren't necessarily in the public or private centers of power are worried that the AI future will look... pretty much exactly like the present, with a lack of coordination in government leading to tech giants riding roughshod over their competitors and notions of privacy or consumer protection.

As outlined in a letter shared with POLITICO's Digital Future Daily yesterday, a group of open-source AI advocates — which, it should be noted, included Meta's AI lead Yann LeCun — are warning that "quickly rushing towards the wrong kind of regulation can lead to concentrations of power in ways that hurt competition and innovation," echoing worries that to restrict AI models' use too tightly will simply mean that a potentially world-changing technology is used to uphold the current status quo.

SUBSCRIBE TO CALIFORNIA CLIMATE: Climate change isn't just about the weather. It's also about how we do business and create new policies, especially in California. So we have something cool for you: A brand-new California Climate newsletter. It's not just climate or science chat, it's your daily cheat sheet to understanding how the legislative landscape around climate change is shaking up industries across the Golden State. Subscribe now to California Climate to keep up with the changes.

A POSITIVE OUTLOOK ON THE METAVERSE

A tech consulting firm's annual trend report is bullish on the metaverse — largely because of the boost it might get from generative AI.

The 2024 tech and media report from the consulting group Activate expects there to be more than 600 million metaverse users globally by 2026, defining "metaverse user" as someone who uses any tool related to a virtual world from fully immersive experiences like Meta's to games like Minecraft or Microsoft's virtual workspace software.

The secret sauce for metaverse growth, they predict: Generative AI tools that will allow those people not just to use virtual spaces, but design them. They note that 67 percent of those who use generative AI to make online content are also active in virtual worlds, and that increasingly sophisticated technology will erase some of the problems with avatars that have made existing metaverse spaces look somewhat low rent.

PRAYING WITH AI

While the rest of the world figures out how to manage an AI-powered future, South Koreans are integrating the technology into a very old practice: Christianity.

The Financial Times recently reported (paywall) on the tens of thousands of South Koreans who are using chatbots and other AI tools as part of their religious practice. (Think of a bot called "Ask Jesus," or one that assists busy pastors in writing sermons.)

"We faced strong resistance from churches initially with their suspicion that we are trying to replace God and pastors," one AI executive told the FT's reporters. "But pastors began to appreciate our service as it helps them save time in preparing for sermons, and find more time to take care of lonely, troubled followers."

The effect might not be entirely salutary, however. When I interviewed the "digital theologian" Adam Graber in July on how he thought AI and Christianity might collide, he had a slightly more jaundiced view of its effect on scriptural interpretation: "There's a semantic overload of meaning... we become overloaded with these different interpretations, and we don't actually know how to make sense of what's a legitimate interpretation or not. Even before AI systems we were struggling with that, and I think AI systems will amplify that even further."

TWEET OF THE DAY

×

THE FUTURE IN 5 LINKS

4

- LinkedIn is the latest platform to add a suite of **OpenAI-assisted** super-tools.

- Tesla **won a lawsuit** that blamed its self-driving software for a deadly crash.

- A **commercial spaceplane** is nearly ready for its first official flight.

- AI might help determine whether **reforestation projects** are actually working.

- Eric Schmidt is betting on AI revolutionizing **scientific research**.

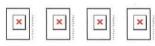*Stay in touch with the whole team: Ben Schreckinger (bschreckinger@politico.com); Derek Robertson (drobertson@politico.com); Mohar Chatterjee (mchatterjee@politico.com); Steve Heuser (sheuser@politico.com); Nate Robson (nrobson@politico.com) and Daniella Cheslow (dcheslow@politico.com).*

If you've had this newsletter forwarded to you, you can sign up and read our mission statement at the links provided.

Follow us on Twitter

Ben Schreckinger @SchreckReports
Derek Robertson @afternoondelete
Steve Heuser @sfheuser

FOLLOW US

s22

From:	s22
Sent:	Thursday, 2 November 2023 7:10 AM
To:	Gale, Brett s22
Subject:	Fwd: Day 1 Evening media wrap [SEC=OFFICIAL]

OFFICIAL

Thought you may be interested....

s22

OFFICIAL

> **CAUTION** - This email originated from outside of the organisation. Do not click links or open attachments unless you recognise the sender and know the content is safe.

OFFICIAL

Evening High Commissioner'

Here is a media round up of articles re: AI Safety Summit at the end of Day 1.

Mainly UK/International but a few Australian stories at the top of the list.

Thanks s22

Global tech rivals pledge to join forces against the machines

Albanese government 'reluctant' to bail out stricken Tritium

UK, US, China sign AI safety pledge at UK summit

US and China join global leaders to lay out need for AI rulemaking

Global Leaders Warn A.I. Could Cause 'Catastrophic' Harm

AI Safety Summit discussions not focused on killer robots, says Darktrace boss

King Charles: Tackle AI risks with urgency and unity

1

Rishi Sunak: AI firms cannot 'mark their own homework

Kamala Harris: AI could enable cyber attacks at scale beyond anything we have ever seen

China will attend Rishi Sunak's AI summit despite backlash from Tory MPs

Elon Musk calls AI 'one of the biggest threats to humanity' at summit

UK, US, EU and China sign declaration of AI's 'catastrophic' danger

The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023

s22

Public Diplomacy
Australia House
Australian High Commission
Strand, London. WC2B 4LA

s22

AUSTRALIA HOUSE
Australia's home in the UK

s22

From:	Murfett, Anthony
Sent:	Tuesday, 31 October 2023 9:28 AM
To:	s22 Quinn, Daniel; Wilson, Helen
Cc:	Gale, Brett; s22
Subject:	RE: WH AI EO [SEC=OFFICIAL]

Hi s22

Thanks for the below. We'll work with Helen to support a briefing.

Cheers

Anthony Murfett
Department of Industry, Science and Resources
s22
s22   @industry.gov.au

Ngunnawal Country, Industry House, 10 Binara Street (GPO Box 2013) Canberra ACT 2601 Australia

industry.gov.au ABN 74 599 608 295
Supporting economic growth and job creation for all Australians | We are collaborative, innovative, respectful and strive for excellence

Acknowledgement of Country
Our department recognises the First Peoples of this nation and their ongoing connection to culture and country. We acknowledge First Nations Peoples as the Traditional Owners, Custodians and Lore Keepers of the world's oldest living culture and pay respects to their Elders past, present and emerging.

OFFICIAL

From: s22   @industry.gov.au>
Sent: Tuesday, 31 October 2023 8:54 AM
To: Quinn, Daniel s.22   @industry.gov.au>; Murfett, Anthony s.22   @industry.gov.au>; Wilson, Helen s.22   @industry.gov.au>
Cc: s.22   @industry.gov.au>
Subject: FW: WH AI EO [SEC=OFFICIAL]

OFFICIAL

Hi all. Not sure who's in the UK with Helen but I've flicked this across to the Minister, via s22

The Minister's asked for a briefing prior to the Summit starting. Suspect its best done verbally by Helen?

s22

1

s22

@industry.gov.au

OFFICIAL

From: s22
Sent: Tuesday, 31 October 2023 8:25 AM
To: s22   @industry.gov.au>
Cc: s.22   @industry.gov.au>
Subject: WH AI EO [SEC=OFFICIAL]

As per text

s22

Senior Advisor
Office of the Hon Ed Husic MP
Minister for Industry and Science
s22
s22   s@industry.gov.au

OFFICIAL

FACT SHEET: President Biden Issues Executive Order
on Safe, Secure, and Trustworthy Artificial Intelligence

Today, President Biden issued a landmark Executive Order to ensure that America leads the way in seizing the promise and managing the risks of artificial intelligence (AI). The Executive Order establishes new standards for AI safety and security, protects Americans' privacy, advances equity and civil rights, stands up for consumers and workers, promotes innovation and competition, advances American leadership around the world, and more.

As part of the Biden-Harris Administration's comprehensive strategy for responsible innovation, the Executive Order builds on previous actions the President has taken, including work that led to voluntary commitments from 15 leading companies to drive safe, secure, and trustworthy development of AI.

The Executive Order directs the following actions:

<u>New Standards for AI Safety and Security</u>
As AI's capabilities grow, so do its implications for Americans' safety and security. **With this Executive Order, the President directs the most sweeping actions ever taken to protect Americans from the potential risks of AI systems:**

- **Require that developers of the most powerful AI systems share their safety test results and other critical information with the U.S. government.** In accordance with the Defense Production Act, the Order will require that companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety must notify the federal government when training the model, and must share the results of all red-team safety tests. These measures will ensure AI systems are safe, secure, and trustworthy before companies make them public.
- **Develop standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy.** The National Institute of Standards and Technology will set the rigorous standards for extensive red-team testing to ensure safety before public release. The Department of Homeland Security will apply those standards to critical infrastructure sectors and establish the AI Safety and Security Board. The Departments of Energy and Homeland Security will also address AI systems' threats to critical infrastructure, as well as chemical, biological, radiological, nuclear, and cybersecurity risks. Together, these are the most significant actions ever taken by any government to advance the field of AI safety.
- **Protect against the risks of using AI to engineer dangerous biological materials** by developing strong new standards for biological synthesis screening. Agencies that fund life-science projects will establish these standards as a condition of federal funding, creating powerful incentives to ensure appropriate screening and manage risks potentially made worse by AI.
- **Protect Americans from AI-enabled fraud and deception by establishing standards and best practices for detecting AI-generated content and authenticating official content.** The Department of Commerce will develop guidance for content authentication and watermarking to

clearly label AI-generated content. Federal agencies will use these tools to make it easy for Americans to know that the communications they receive from their government are authentic—and set an example for the private sector and governments around the world.

- **Establish an advanced cybersecurity program to develop AI tools to find and fix vulnerabilities in critical software,** building on the Biden-Harris Administration's ongoing AI Cyber Challenge. Together, these efforts will harness AI's potentially game-changing cyber capabilities to make software and networks more secure.
- **Order the development of a National Security Memorandum that directs further actions on AI and security,** to be developed by the National Security Council and White House Chief of Staff. This document will ensure that the United States military and intelligence community use AI safely, ethically, and effectively in their mission, and will direct actions to counter adversaries' military use of AI.

Protecting Americans' Privacy

Without safeguards, AI can put Americans' privacy further at risk. AI not only makes it easier to extract, identify, and exploit personal data, but it also heightens incentives to do so because companies use data to train AI systems. **To better protect Americans' privacy, including from the risks posed by AI, the President calls on Congress to pass bipartisan data privacy legislation to protect all Americans, especially kids, and directs the following actions:**

- **Protect Americans' privacy by prioritizing federal support for accelerating the development and use of privacy-preserving techniques**—including ones that use cutting-edge AI and that let AI systems be trained while preserving the privacy of the training data.
- **Strengthen privacy-preserving research and technologies,** such as cryptographic tools that preserve individuals' privacy, by funding a Research Coordination Network to advance rapid breakthroughs and development. The National Science Foundation will also work with this network to promote the adoption of leading-edge privacy-preserving technologies by federal agencies.
- **Evaluate how agencies collect and use commercially available information**—including information they procure from data brokers—and **strengthen privacy guidance for federal agencies** to account for AI risks. This work will focus in particular on commercially available information containing personally identifiable data.
- **Develop guidelines for federal agencies to evaluate the effectiveness of privacy-preserving techniques,** including those used in AI systems. These guidelines will advance agency efforts to protect Americans' data.

Advancing Equity and Civil Rights

Irresponsible uses of AI can lead to and deepen discrimination, bias, and other abuses in justice, healthcare, and housing. The Biden-Harris Administration has already taken action by publishing the Blueprint for an AI Bill of Rights and issuing an Executive Order directing agencies to combat algorithmic discrimination, while enforcing existing authorities to protect people's rights and safety. **To ensure that AI advances equity and civil rights, the President directs the following additional actions:**

- Provide clear guidance to landlords, Federal benefits programs, and federal contractors to keep AI algorithms from being used to exacerbate discrimination.
- Address algorithmic discrimination through training, technical assistance, and coordination between the Department of Justice and Federal civil rights offices on best practices for investigating and prosecuting civil rights violations related to AI.
- Ensure fairness throughout the criminal justice system by developing best practices on the use of AI in sentencing, parole and probation, pretrial release and detention, risk assessments, surveillance, crime forecasting and predictive policing, and forensic analysis.

Standing Up for Consumers, Patients, and Students

AI can bring real benefits to consumers—for example, by making products better, cheaper, and more widely available. But AI also raises the risk of injuring, misleading, or otherwise harming Americans. **To protect consumers while ensuring that AI can make Americans better off, the President directs the following actions:**

- Advance the responsible use of AI in healthcare and the development of affordable and life-saving drugs. The Department of Health and Human Services will also establish a safety program to receive reports of—and act to remedy – harms or unsafe healthcare practices involving AI.
- Shape AI's potential to transform education by creating resources to support educators deploying AI-enabled educational tools, such as personalized tutoring in schools.

Supporting Workers

AI is changing America's jobs and workplaces, offering both the promise of improved productivity but also the dangers of increased workplace surveillance, bias, and job displacement. **To mitigate these risks, support workers' ability to bargain collectively, and invest in workforce training and development that is accessible to all, the President directs the following actions:**

- Develop principles and best practices to mitigate the harms and maximize the benefits of AI for workers by addressing job displacement; labor standards; workplace equity, health, and safety; and data collection. These principles and best practices will benefit workers by providing guidance to prevent employers from undercompensating workers, evaluating job applications unfairly, or impinging on workers' ability to organize.
- Produce a report on AI's potential labor-market impacts, and study and identify options for strengthening federal support for workers facing labor disruptions, including from AI.

Promoting Innovation and Competition

America already leads in AI innovation—more AI startups raised first-time capital in the United States last year than in the next seven countries combined. **The Executive Order ensures that we continue to lead the way in innovation and competition through the following actions:**

- **Catalyze AI research across the United States** through a pilot of the National AI Research Resource—a tool that will provide AI researchers and students access to key AI resources and data—and expanded grants for AI research in vital areas like healthcare and climate change.
- **Promote a fair, open, and competitive AI ecosystem** by providing small developers and entrepreneurs access to technical assistance and resources, helping small businesses commercialize AI breakthroughs, and encouraging the Federal Trade Commission to exercise its authorities.
- **Use existing authorities to expand the ability of highly skilled immigrants and nonimmigrants with expertise in critical areas to study, stay, and work in the United States** by modernizing and streamlining visa criteria, interviews, and reviews.

Advancing American Leadership Abroad

AI's challenges and opportunities are global. **The Biden-Harris Administration will continue working with other nations to support safe, secure, and trustworthy deployment and use of AI worldwide. To that end, the President directs the following actions:**

- **Expand bilateral, multilateral, and multistakeholder engagements to collaborate on AI.** The State Department in collaboration with the Commerce Department will lead an effort to establish robust international frameworks for harnessing AI's benefits and managing its risks and ensuring safety. In addition, this week, Vice President Harris will speak at the UK Summit on AI Safety, hosted by Prime Minister Rishi Sunak.
- **Accelerate development and implementation of vital AI standards** with international partners and in standards organizations, ensuring that the technology is safe, secure, trustworthy, and interoperable.
- **Promote the safe, responsible, and rights-affirming development and deployment of AI abroad to solve global challenges,** such as advancing sustainable development and mitigating dangers to critical infrastructure.

Ensuring Responsible and Effective Government Use of AI

AI can help government deliver better results for the American people. It can expand agencies' capacity to regulate, govern, and disburse benefits, and it can cut costs and enhance the security of government systems. However, use of AI can pose risks, such as discrimination and unsafe decisions. **To ensure the responsible government deployment of AI and modernize federal AI infrastructure, the President directs the following actions:**

- **Issue guidance for agencies' use of AI,** including clear standards to protect rights and safety, improve AI procurement, and strengthen AI deployment.
- **Help agencies acquire specified AI products and services** faster, more cheaply, and more effectively through more rapid and efficient contracting.
- **Accelerate the rapid hiring of AI professionals** as part of a government-wide AI talent surge led by the Office of Personnel Management, U.S. Digital Service, U.S. Digital Corps, and Presidential

Innovation Fellowship. Agencies will provide AI training for employees at all levels in relevant fields.

As we advance this agenda at home, the Administration will work with allies and partners abroad on a strong international framework to govern the development and use of AI. The Administration has already consulted widely on AI governance frameworks over the past several months—engaging with Australia, Brazil, Canada, Chile, the European Union, France, Germany, India, Israel, Italy, Japan, Kenya, Mexico, the Netherlands, New Zealand, Nigeria, the Philippines, Singapore, South Korea, the UAE, and the UK. The actions taken today support and complement Japan's leadership of the G-7 Hiroshima Process, the UK Summit on AI Safety, India's leadership as Chair of the Global Partnership on AI, and ongoing discussions at the United Nations.

The actions that President Biden directed today are vital steps forward in the U.S.'s approach on safe, secure, and trustworthy AI. More action will be required, and the Administration will continue to work with Congress to pursue bipartisan legislation to help America lead the way in responsible innovation.

For more on the Biden-Harris Administration's work to advance AI, and for opportunities to join the Federal AI workforce, visit AI.gov.