
CAIDP Update 6.33

September 2, 2024

Welcome to this week's CAIDP Update on global AI policy and governance developments. As we enter September, the AI landscape continues to evolve with significant policy shifts.

This week, the Council of Europe's AI Convention opens for signatures on September 5, 2024. In the U.S., NIST's AI Safety Institute has partnered with Anthropic and OpenAI for research, while California's AI safety bill, SB 1047, awaits the governor's approval. Additionally, the FTC has launched a new rule to tackle AI-generated fake reviews.

Russia and China agree to boost AI cooperation, while Colombia establishes a commission for AI regulation. Denmark rules on AI-powered call analysis in customer service, and Bosnia's data protection agency warns about Meta's privacy policy changes affecting non-EU users.

On the advocacy front, CAIDP Europe backs the Swiss campaign for "Artificial Intelligence Without Discrimination," and CAIDP's president discusses global AI governance with the UN Under-Secretary-General. The Center continues to push for amendments to the UN Global Digital Compact.

AI POLICY NEWS

First International AI Treaty Opens for Signature

The Council of Europe's [Convention on AI, Human Rights, Democracy, and the Rule of Law](#), the world's first binding international treaty on AI, will open for signature on September 5, 2024, in [Vilnius](#). This landmark [treaty](#) establishes fundamental principles for AI systems, including "Human dignity and individual autonomy, Equality and non-discrimination, Respect for privacy and personal data protection."



The Convention requires signatories to "carry out risk and impact assessments in respect of actual and potential impacts on human rights, democracy and the rule of law." The Convention covers "the activities within the lifecycle of artificial intelligence systems undertaken by public authorities, or private actors acting on their behalf."

The European Council has also [authorized](#) the EU's signing of the Convention. The

Council Decision notes, "Activities within the lifecycle of AI systems related to the protection of national security interests are excluded from the scope of the Convention." However, the Convention states that "parties must ensure that such activities respect international law and democratic institutions and processes."

Numerous civil society organizations including CAIDP were actively involved in shaping the AI Treaty. While the treaty did not fully meet our expectations, particularly in its lack of comprehensive regulations for private-sector AI systems, we strongly advocate for the endorsement of the Council of Europe's AI Treaty. For more information, visit the CAIDP resource page [here](#).

NIST AI Safety Accord Emerges as California's AI Bill Reaches Governor's Desk

The U.S. AI Safety Institute at NIST signed first-of-their-kind [agreements](#) with Anthropic and OpenAI for AI safety research collaboration. This federal



initiative establishes a framework for NIST to receive access to major new AI models from both companies before and after their public release.

Meanwhile, California's [Senate Bill 1047](#), requiring safety protocols for large AI models, has passed the legislature and now awaits Governor Gavin Newsom's decision. OpenAI [opposes](#) the bill, arguing for federal-level regulation to address national security concerns and maintain state competitiveness.

Anthropic, however, cautiously [supports](#) SB 1047, believing its benefits outweigh costs after recent amendments. They stress the importance of addressing catastrophic AI risks but raise concerns about certain aspects, including pre-harm enforcement mechanisms.

The timing of these agreements raises questions about the companies' stances on AI governance. The willingness to cooperate with voluntary federal initiatives contrasts with the varied responses to state-level regulation. It also prompts inquiry into whether federal collaboration reflects a genuine commitment to safety or a strategic move to influence the regulatory landscape.

Despite industry division, the bill has gained significant support. AI experts like Geoffrey Hinton and Yoshua Bengio endorse it, while organizations like Encode Justice and CAIDP argue it addresses crucial regulatory gaps. Elon Musk also stated on X that California should "probably pass the SB 1047 AI safety bill."

FTC Unveils Rule to Combat Fake Reviews

The Federal Trade Commission (FTC) issued a [final rule](#) to tackle fake reviews and testimonials, bolstering its enforcement capabilities. The measure prohibits the sale, purchase, and creation of false consumer feedback, including AI-generated content.



"Fake reviews not only waste people's time and money, but also pollute the marketplace and divert business away from honest competitors," said FTC Chair Lina M. Khan.

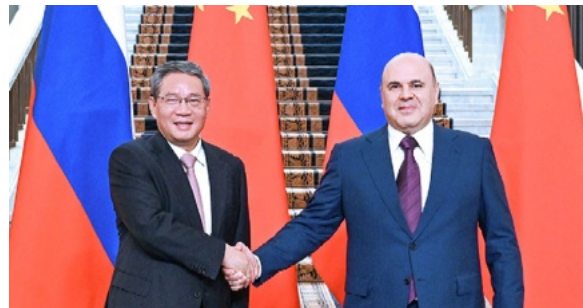
The rule bans "buying positive or negative reviews" and "review suppression." It also addresses "insider reviews" and "company-controlled review websites."

"By strengthening the FTC's toolkit to fight deceptive advertising, the final rule will protect Americans from getting cheated," Khan added.

Set to take effect 60 days after Federal Register publication, the rule aims to "enhance deterrence and strengthen FTC enforcement actions" in light of recent legal challenges to its authority.

Russia and China Agree to Boost AI Cooperation

Russia and China agreed to strengthen cooperation in AI, as announced in a joint communiqué following a meeting between Russian Prime Minister Mikhail Mishustin and Chinese Premier Li Qiang. The nations plan to establish a working group on AI cooperation within an existing bilateral commission.



"They agreed to intensify practical cooperation in the field of artificial intelligence on the basis of mutual respect, equality and mutual benefit," the document states. The focus will be on "global governance" and "the regulation of ethics and industrial application of artificial intelligence."

The nations also pledged to support AI capacity building among BRICS (Brazil, Russia, India, China, and South Africa) countries. China will also establish a "China-BRICS Artificial Intelligence Development Center" to promote "industry connectivity and capacity building."

Danish Agency Rules on IDA Forsikring's AI Call Analysis

The Danish Data Protection Agency concluded its investigation into IDA Forsikring's use of AI to analyze recorded customer service calls.



While the agency found that IDA Forsikring can record and analyze calls within data protection rules, it determined that the company's current consent process is inadequate.

The investigation examined IDA Forsikring's practice of recording incoming calls and using AI to convert audio files to text for analysis. The agency stated that IDA Forsikring's consent process "does not comply with data protection rules," as it fails to offer callers separate consent options for different processing purposes.

The agency did not criticize IDA Forsikring but advised the company to ensure its future processes align with data protection regulations. The decision highlights the importance of proper consent mechanisms in AI-driven data processing.

Colombia Sets Up Commission for AI Regulation

Colombia's Congress created a new Commission to coordinate and refine AI regulation. Announced on August 27, 2024, the commission's primary task is to consolidate nine existing legislative initiatives aimed at ensuring the ethical development of AI technology.



The establishment of the commission is a move to create a cohesive legislative framework that adheres to the principles of transparency, equity, and justice. This aligns various bills into a unified policy for AI in Colombia.

The government also plans to build AI research centers and implement educational programs to support the development and application of AI technologies throughout the nation. These steps are integral to the commission's broader mandate to oversee AI legislation.

Caution Advised for Meta Users in Bosnia and Herzegovina

The Agency for Personal Data Protection in Bosnia and Herzegovina issued a warning to users of Meta Platforms, which include Facebook, Instagram, and WhatsApp. The agency has expressed concerns regarding recent unannounced changes to the privacy policy by Meta that affect users in Bosnia and Herzegovina, as well as other non-EU countries in the region.



On June 26, 2024, Meta modified its privacy policy to expand the use of user data including personal information, photos, videos, posts, and comments for the development of AI technologies, without clear communication of the purposes or methods involved. Unlike EU citizens, who were notified and given the option to opt out, users in Bosnia and Herzegovina received no such information or choice.

The lack of transparency and the inability to refuse data usage have raised significant concerns about the protection of personal data and privacy rights under the new policy. The agency has already contacted Meta to request the appointment of a representative to address these issues but has not yet received a response.

Join the movement & empower the AI policy change-makers

Your contribution makes a real difference:

Provide invaluable AI policy training to aspiring leaders in over 80 countries.

Support the development of our AI & Democratic Values Index Report, a vital resource for understanding AI's global impact.

Stay informed with our comprehensive updates on AI policy developments worldwide.

Help us continue our essential dialogue with policymakers, shaping the responsible evolution of AI.

Every donation helps the Center for AI and Digital Policy build a more informed and equitable AI-driven world.

[Donate with PayPal](#)

CAIDP ACTIONS

CAIDP Europe Supports Swiss Campaign for Equitable AI

CAIDP Europe, the European branch of the CAIDP, [joined](#) the Swiss campaign for "Artificial Intelligence Without Discrimination." This collaboration underscores a shared commitment to ensuring that AI systems promote fairness and do not perpetuate biases.



Systemic discrimination via AI has been increasingly evident, as outlined in CAIDP's [AI & Democratic Values Index](#), which monitors AI policies in 80 countries. Notably, incidents such as the Dutch childcare benefit scandal and the British Post Office scandal have illuminated the severe impacts of AI-driven discrimination, including racial profiling and faulty system designs leading to wrongful accusations and severe personal and financial hardships.

As for Switzerland, discrimination based on AI reproduction of gender or origin bias, is of particular concern. A Swiss study has demonstrated that these are the main sources of discrimination in recruitment. Trust in AI has also decreased in the Swiss population. In 2023, a survey revealed that 60% of the population had no trust at all in, or did not really trust, the use of AI tools.

The Swiss government should correct existing gaps in the legislation to ensure respect for these principles and empower Swiss citizens in the digital era. This is the reason why CAIDP Europe is joining AlgorithmWatch Switzerland and 43 other organisations in urging the Swiss government to uphold "Artificial Intelligence Without Discrimination".

CAIDP Europe calls on other civil society organisations, academics, researchers and citizens to join this important initiative. To participate, please click [here](#). The call is open for signature until 23 September 2024.

Secretary-General Marwala

This week, Merve Hickok, President of the CAIDP, met with Tshilidzi Marwala, the Under-Secretary-General of the United Nations and Rector of the United Nations University, in Tokyo. The meeting, part of Hickok's fellowship with the Council on Foreign Relations - Hitachi International Affairs Fellowship and her collaborative work with Chiba Institute of Technology, facilitated discussions on global AI governance.



Joined by Chief of Staff Michael Baldock, the dialogue covered a range of topics: recent Tokyo International Conference on African Development (TICAD) ministerial meetings, Data Free Flow with Trust, the UN's expanding role in AI governance, and the forthcoming Summit for the Future and Global Digital Compact. Discussions also explored the role of incentives in shaping behaviors across human, corporate, and national spectrums.

Update on the UN Global Digital Compact (GDC)

On another front, the latest draft (rev4) of the UN Global Digital Compact was published online on August 27, 2024.

Launched by UN Secretary-General António Guterres in 2021, the GDC aims to delineate shared principles for an open, free, and secure digital future, with expected adoption at the September 2024 Summit for the Future.



In its ongoing advocacy, CAIDP has urged for specific amendments to the GDC: the establishment of a UN Special Rapporteur for AI and Human Rights, creation of a Scientific Panel on AI, a ban on Lethal Autonomous Weapons Systems (LAWS), and the adoption of the Universal Guidelines for AI as foundational for global governance.

CAIDP has strongly advocated for the UN to encourage member states to adopt transparent AI governance processes and to assess the impact of AI on fundamental rights detailed in the International Covenant on Civil and Political Rights and the Universal Declaration of Human Rights. These rights include self-determination and privacy among others.

In drafting guidance, CAIDP has highlighted the need to clarify that "interoperable" refers only to technical standards, not governance frameworks or fundamental rights, and to avoid using the term "platforms" for internet companies. It has also called for stringent prohibitions on AI deployments that fail to meet international human rights standards, reinforcing its commitment to a human-centric approach to AI governance.

AI POLICY EVENTS

- [National Academies Workshop on “Emerging technologies to aid behavioral](#)

threat assessment with Christabel Randolph (CAIDP), Washington DC, September 3, 2024

- Council of Europe AI Treaty Open for Ratification, Vilnius, September 5, 2024
- Design It For Us with Marc Rotenberg (CAIDP), American University, Washington DC, September 11, 2024
- AI Law and Governance, International Bar Association with Marc Rotenberg (CAIDP), Mexico City, September 19, 2024
- Unlocking Cooperation: AI for All with Christabel Randolph (CAIDP), Carnegie Council and UNU Centre for Policy Research, September 19, 2024
- 11th Plenary meeting of the Committee on AI, Council of Europe, Strasbourg, France, September 17-19, 2024
- AI Policy Forum with Christabel Randolph (CAIDP), Partnership on AI, New York, September 20, 2024
- Summit of the Future / Global Digital Compact with Ayca Ariyörük (CAIDP), United Nations, New York, September 20-23, 2024
- CONVERGE 2024: The Indo-Pacific Critical Technologies Forum with Merve Hickok (CAIDP), Tokyo, October 10-11, 2024
- Annual Asia Privacy Bridge Forum with Merve Hickok (CAIDP), Seoul, October 17, 2024
- AI and Repression: Emerging Challenges for Dissidents and Journalists, Frontlines of Freedom, Renew Democracy Initiative and the CAIDP, Washington DC, October 29-30, 2024
- AI and Governance: Addressing ethical, legal and operational challenges with Varsha Sewlal (CAIDP), Johannesburg, October 29, 2024
- Global HR Forum 2024 with Merve Hickok (CAIDP), Seoul, October 30-31, 2024
- After the US Election: The Future of the AI Executive Order and US AI Policy” with Marc Rotenberg (CAIDP), Université du Luxembourg, November 7, 2024
- Women & Justice Summit with Merve Hickok (CAIDP), Istanbul, November 8, 2024
- AI in Global Society with Marc Rotenberg (CAIDP), Harvard Business School, Cambridge, November 20, 2024

[Subscribe to CAIDP](#) | Let us know your suggestions [here](#).

| [Privacy Policy](#) | [Contact](#) | [Donate](#) |



CAIDP | 1100 13th ST NW Suite 800 | Washington, DC 20005 US

[Unsubscribe](#) | [Update Profile](#) | [Constant Contact Data Notice](#)



Try email marketing for free today!