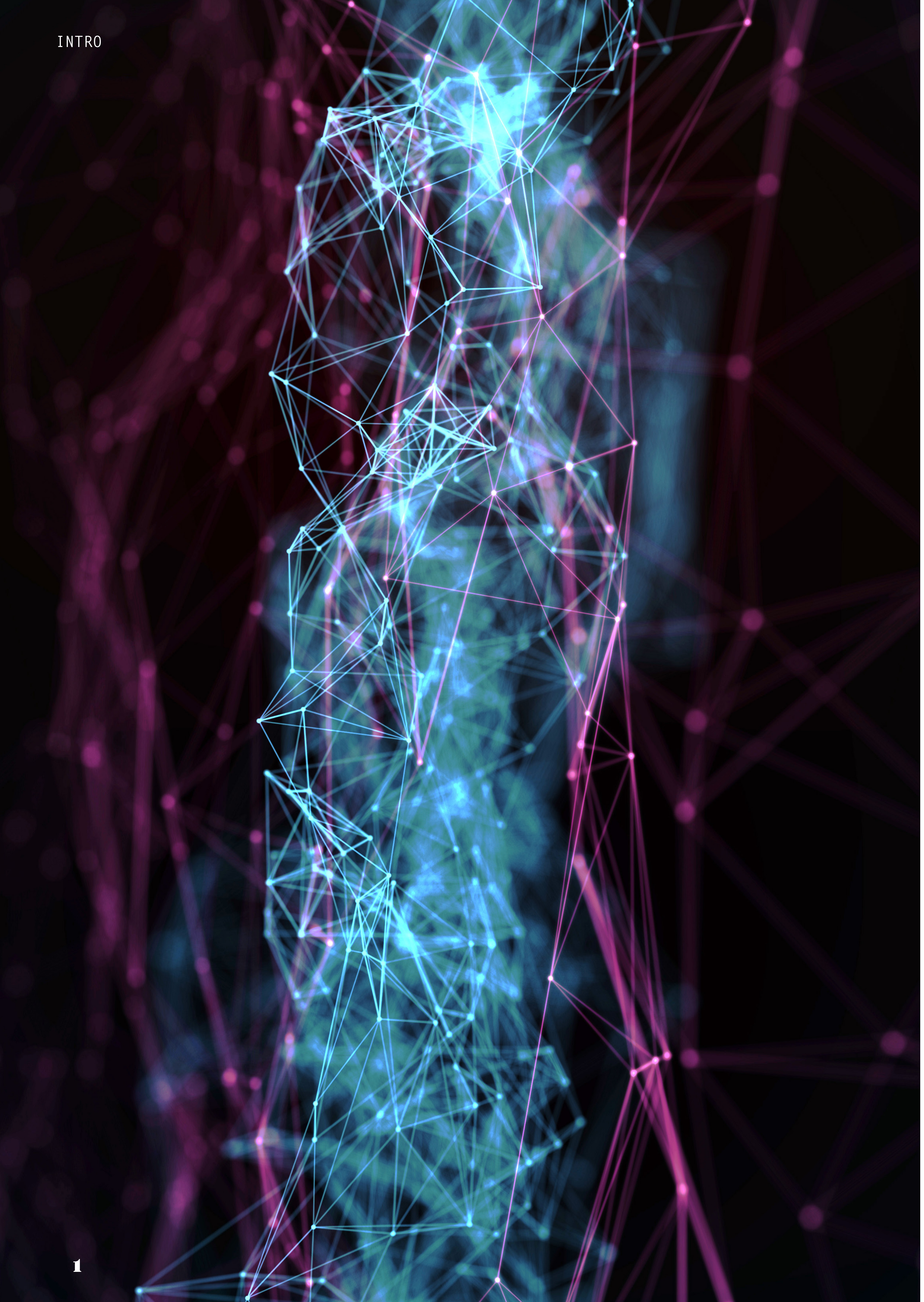
Bertelsmann
FOUNDATION

HIDDEN LAYERS

YOUR QUARTERLY TRANSATLANTIC TECHNOLOGY NEWS

November 2023



Welcome back to another edition of Hidden Layers. In this issue, we discuss recent developments in artificial intelligence (AI) regulation: the Biden administration's executive order (EO) on AI, the UK AI Safety Summit, the EU's AI Act, and an executive order to regulate generative AI in California.

REGULATING AI

Recent technology regulation has emphasized AI governance, as leading political powers race to become the first to regulate the technology and set global standards. In the U.S., the Biden administration issued a long-awaited EO on AI, the EU entered trilogue negotiations on the AI Act, and the UK brought together high-level government officials and technology experts at Bletchley Park to discuss future global collaboration on AI.

AT THE WHITE HOUSE

The Biden administration continues to make a concerted effort to address AI-related risks and build a comprehensive strategy for responsible innovation. In addition to securing the voluntary commitments on AI announced this summer, the president signed on October 30 a landmark EO on safe, secure and trustworthy AI that seeks to balance harnessing its power with mitigating its substantial risks.

The EO provides eight guiding principles for its ambitious recommendations and actions: new standards for AI safety and security; protecting Americans' privacy; advancing equity and civil rights; protecting consumers, patients and students; supporting workers; promoting innovation and competition; advancing American leadership abroad; and ensuring responsible and effective government use of AI.

The EO is an important first step to creating guardrails for AI developers that ensure the technology is deployed in a way that guarantees consumer trust and safety without stifling corporate innovation and profitability. The order is also comprehensive and adopts a whole-of-government approach. The National Institute of Standards and Technology (NIST) at the Department of Commerce will oversee the development of AI guidelines. Technology companies can use NIST's AI Risk Management Framework as a baseline for their work, and the agency will set rigorous standards for red-team testing to ensure the safety of AI systems before their public release. The Department of Homeland Security will apply NIST's standards to critical infrastructure sectors and, along with the Department of Energy, address threats from AI systems to those sectors, particularly the chemical, biological, nuclear and cybersecurity risks. The Department of Justice will address algorithmic discrimination and coordinate with federal civil rights offices on investigating and prosecuting AI-related civil rights violations, while the Department of Health and Human Services will tackle any relevant unsafe healthcare practices.

> *"This EO is the 'bat signal' that the global AI Policy ecosystem needed - it signals the U.S. government's intention to lead in developing responsible AI, with almost every U.S. government department and agency tasked to develop Responsible AI practices within a compressed period. The effects will cascade to many sectors of the economy, and affect AI governance worldwide," notes Evi Fuelle of Credo AI, a startup focused on responsible AI and governance, who attended the White House signing of the EO.*

Although the EO is the most significant U.S. action taken to date to regulate AI at the federal level, industry and civil society have expressed concern about some aspects of the EO. Those who view AI's potential optimistically believe the order will hinder innovation, an argument that has been used repeatedly to oppose technology regulation. Civil society organizations see the EO as a meaningful first step, but one that does not go far enough to ensure AI fairness or address dangers that AI models pose, especially discrimination against marginalized communities. The order also does not mention comprehensive data privacy legislation, which Congress has not passed. Large amounts of data are necessary to develop AI models, and experts argue that AI systems could reveal sensitive and confidential personal information in the absence of strong data privacy laws.

The technology sector is also concerned about developments on the international level. They want global interoperability to avoid having to comply with a jumble of national AI guidelines and regulations, some of which could compete with or contradict one another.

DISCUSSIONS ON CAPITOL HILL

To be effective, Biden's EO likely needs Congress to pass parallel legislation, but lawmakers remain unsure about the best way to mitigate AI-related risks. Lawmakers in both chambers held hearings on AI in early November to increase their understanding of the issues at hand. The Senate Committee on Homeland Security and Governmental Affairs explored the philosophy and history of AI and its lessons for regulation in today's digital environment. The body's Health, Education, Labor and Pensions Subcommittee examined AI's impact on health care access given algorithmic bias in systems that private insurers use to approve or deny claims. The House Committee on Oversight and Accountability looked into advances in deepfakes and their capacity to be weaponized, an especially acute issue given next year's presidential election and the dangers of disinformation.

While Congress is having these discussions at the federal level, state regulators have already taken action to regulate AI. California Governor Gavin Newsom signed in September an EO that regulates the development and use of generative AI (GenAI). Since the state is home to 35 of the world's top 50 AI companies and has become a global hub for the technology, California legislators feel a need to be at the forefront of its regulation. The EO directs state agencies to perform a risk analysis of potential threats to and vulnerabilities of critical energy infrastructure caused by GenAI. The order also provides a blueprint for public sector procurement of GenAI applications, and a deployment and analysis framework that develops guidelines for agencies to analyze the impact that adopting GenAI tools may have on vulnerable communities. State employees will receive training to develop skills needed to thrive in an AI economy. The state senate, for its part, is debating a bill that proposes a safety framework for AI to promote the safe and secure development of large-scale AI systems.

If you are interested in the role of deepfakes and generative AI in the 2024 U.S. elections, look out for "Deep Fakes and Deep Trouble: The Political Consequences of AI-Generated Ads" in the next issue of Transponder, BFNA's biannual publication on issues impacting the transatlantic relationship.

A GATHERING AT BLETCHLEY PARK

Two days after Biden issued his EO, on November 1-2, global political leaders and AI experts, including those from the U.S., EU and China, met in the UK for the first AI Safety Summit. The event was an opportunity for the British to show global leadership on AI governance and establish themselves as a nexus among the world's three largest economic blocs on this issue. The UK met its main objectives for the summit, which included developing a shared understanding of AI-related risks, considering the ability to mitigate them through coordinated international action, and establishing a global forum for further discussions on AI.

The summit concluded with the signing, by 28 countries, of the Bletchley Declaration, which recognizes the need for more research and coordination to seize the opportunities that AI offers but ensure its safe development. The declaration also included language about AI's greatest risks, particularly in the areas of cybersecurity, biotechnology and disinformation. Participants committed themselves to meeting in person again in 2024. South Korea will host a virtual gathering in six months.

The summit also led to the creation of a UK-based AI Safety Institute, whose U.S. counterpart will be under NIST's leadership. The new agency will test the safety of emerging AI technologies before and after they are released, conduct research on AI safety, and share information with governments, the private sector and civil society. Several leading AI developers, including Google's DeepMind, have already agreed to partner on testing.

MEANWHILE IN BRUSSELS

After two years of internal discussions on the EU's AI Act, the European Parliament, the European Council and the European Commission started trilogue negotiations this summer to finalize the legislation. The AI Act follows a risk-based approach, meaning it outlines rules for the use of AI systems according to the level of risk that they pose to citizens. AI systems with an "unacceptable risk" would be prohibited in the European market, while those classified as "high risk" would be subject to strict obligations. AI systems of "minimal risk" would be allowed if they follow transparency guidelines.

The classifications seem straightforward, but there has been disagreement within the EU on the systems that fall into each category. The European Parliament wants a broad list of prohibited systems that includes, for example, those that use facial recognition and biometric surveillance in public spaces. The European Council aims for a narrower list with exemptions for high-risk systems used for national security and law enforcement.

At the most recent trilogue meeting that took place on October 24, policymakers agreed on wording to Article 6 of the legislation, which outlines classification rules for high- risk AI systems. The bill's original draft automatically classified as high risk all AI systems included in a previously compiled list of critical-use cases, but the negotiations included discussions on exemptions. The bill's most recent version states that an AI system is not considered high risk if it meets the following conditions: a) it performs a narrow procedural task, such as a system that classifies documents or sifts through datasets; b) it detects deviations from decision-making patterns and flags inconsistencies; c) it only improves the quality of existing work, such as checking grammar; and d) does not influence the outcome of decision-making, for example allowing an AI model to make a decision for a bank loan without human assessment.

EU policymakers also discussed guidance for the use of foundation models, such as Open AI's GPT-4. Following the October trilogue, there seemed to be some consensus on a tiered approach that implements horizontal rules for all foundational models, including transparency guidelines that align with the rest of the legislation, and create additional rules for high-impact foundation models. It is unclear what models would classify as high impact, but policymakers were considering factors such as computing power, the amount of training data used, and economic resources of the AI developers.

The AI Act was on track to be approved by the end of this year, provided that the three European bodies could reach an agreement at their next trilogue meeting on December 6. But a document circulated by France, Germany, and Italy on November 19 is likely to stall negotiations. The three largest economies in the bloc state in the document that they are opposed to the tiered approach for foundation models arguing that it goes against the AI Act's technology-neutral and risk-based approach. They propose that instead of the horizontal rules, the EU should require AI companies developing foundation models to self-regulate by signing up to codes of conduct and publishing information about their models, including capabilities, intended uses, potential limitations, and results of studies on biases. In this proposal, there would initially be no sanctions regime, unless companies repeatedly violate the codes of conduct.

ON PRIVACY

In the last issue of Hidden Layers, we discussed the European Commission's adequacy decision on the Transatlantic Data Privacy Framework (DPF). It restored the free flow of data across the Atlantic, which had been on hold since the European Court of Justice struck down in 2020 the previous framework, the Privacy Shield agreement. The DPF, however, pertains only to data shared between the U.S. and EU member states. Where does that leave the UK?

Since it decided to retain the EU's GDPR rules after Brexit, the UK is under a new but almost identical law, the UK GDPR. The legislation restricts international transfers of personal data outside of the UK, as the EU does. That forced the country's Department of Science and Innovation to follow the bloc and negotiate its own data-sharing agreement with the U.S.

Over the summer, Biden and Prime Minister Rishi Sunak announced an agreement in principle to establish a U.S.-UK Data Bridge, which would serve as an extension of the U.S.-EU DPF. The agreement required the UK to affirm that the U.S. provides a sufficient level of data protection, one comparable to those offered to British citizens under the UK GDPR. In return, the U.S. included the UK as a qualifying state under an EO issued to assuage Europeans' concerns that U.S. signals intelligence activities would not violate EU residents' data privacy.

London followed through on its commitment in September, and its adequacy decision entered into force a month later. British business has welcomed the Data Bridge because it allows for the free flow of personal data between the two countries and the EU, but some criticism about insufficient data protections has arisen. The Data Bridge agreement's definition of sensitive data in is broad and does not match that in the UK GDPR, which requires such data to be categorized (e.g., biometric or genetic). Differences in the U.S. and UK handling of criminal offense data also exist, and it is unclear how such information from the UK is protected once transferred to the U.S.