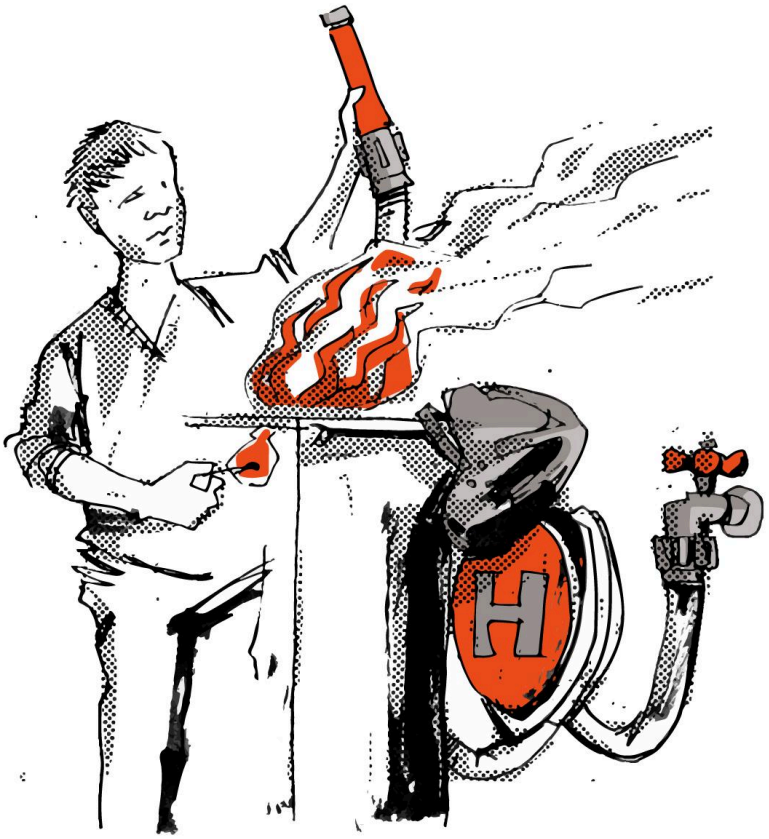
GENEVA INTERNET PLATFORM
digwatch

NEWSLETTER

Issue 86, **February 2024**



AI gurus: between pyromaniacs and firefighters

Snapshot: The developments that made waves

AI governance

The UN Secretary-General's AI Advisory Body launched its Interim Report on governing AI for humanity, while the UN General Assembly adopted a resolution on Lethal Autonomous Weapons Systems (LAWS). Global leaders pledged to promote responsible AI at the 2023 Global Partnership on Artificial Intelligence (GPAI) Summit in New Delhi. Italy took over the G7 presidency in January 2024 and outlined priorities that include Africa's development and AI. The international standard ISO/IEC 42001 on AI management systems was published.

The EU ambassadors have given the green light to the EU AI Act, which is now expected to be formally adopted by the EU by mid-April. The leaked consolidated text of the EU AI Act prompted discussions. The European Commission is preparing to establish the European AI Office, which will be crucial in enforcing the AI Act. The Commission is also challenging the US-led bid to exclude the private sector from the Council of Europe's upcoming Convention on AI and human rights.

The US Federal Trade Commission launched an inquiry into tech giants' AI investments. The National AI Research Resource was launched with to support responsible AI research and drive innovation. The White House released a fact sheet outlining key AI actions following Biden's executive order on AI.

The USA, the UK, and the EU are scrutinising Microsoft's partnership with OpenAI.

Other jurisdictions are active in the AI field as well: Australia is planning to establish an advisory body for AI oversight and regulation, India is entering the LLM race with a Telugu model, China approved over 40 AI models in the last six months, and Saudi Arabia launched the GenAI for All initiative. OpenAI's ChatGPT again faces scrutiny from the Italian data protection authority.

Leading US healthcare companies committed to the ethical adoption of AI, Microsoft pledged to enhance trust and safeguard privacy in the age of AI. OpenAI removed an explicit ban on military use, addressed concerns over election misuse of AI, and unveiled an AI safety framework.

Technologies

Elon Musk's Neuralink implanted the first-ever brain chip in a human. Investigation revealed that the Chinese military bypassed US restrictions on Nvidia chips. Nvidia will roll out a new AI chip for China to comply with US export restrictions. New Dutch restrictions impacted ASML shipments to China. South Korea will extend tax benefits for investments in the nation's semiconductor industry.

Security

Big Tech CEOs testified at a US Senate hearing over accusations of failing to take adequate measures to protect children from harmful and CSAM content.

The OEWG on ICT security issued a call to member states to nominate points of contact to be included in the global points of contact directory on ICT security.

Ukraine's largest telecom operator suffered a crippling cyberattack, allegedly by Russian hackers. The FBI and DoJ used a court order to thwart Chinese hacking of critical infrastructure. A supermassive Mother of all Breaches (MOAB) exposed over 26 billion data records. The UK NCSC warned of the escalating frequency and impact of cyberattacks due to AI. Researchers predicted that cybercrime will incur a staggering USD12 trillion in costs by 2025.

Infrastructure

Negotiations on the EU regulation to accelerate 5G and fibre rollout are underway. Key discussions revolve around the tacit approval principle and intra-EU communication fees.

Legal

The *Times* filed a lawsuit against OpenAI and Microsoft in the USA, alleging that the tech firms used the newspaper's content without authorisation to train their AI large language models. OpenAI responded that the case is 'without merit' and expressed hope for a

partnership with the media outlet. Epic Games has won an antitrust lawsuit against Google over its Play Store app. The IMF chief called for retraining and safety nets amidst AI-driven job changes.

Internet economy

China revealed its plan for a better digital economy and common prosperity, which aims to integrate digital technologies with the real economy and address development issues using digital means. The Association of Southeast Asian Nations (ASEAN) is gearing up to establish a region-wide digital economy agreement. At the same time, the UK's Competition and Markets Authority (CMA) has outlined its plans for implementing a new digital markets competition regime.

Digital rights

A legal filing by the state of New Mexico, the USA, alleges that Meta profits from corporate ads placed alongside content promoting child sexual exploitation. The French Data Protection Authority imposed a €10 million fine against Yahoo for failing to comply with users' privacy and consent regarding cookies.

Content policy

The Cyberspace Administration of China (CAC) focused on curbing pessimism and extremism on digital platforms, while the European Commission started investigating X for alleged breaches of the Digital Services Act (DSA). Meta's Oversight Board criticised Meta for the removal of videos depicting the Israel-Hamas conflict. Türkiye's constitutional court ruled internet content-blocking provisions unconstitutional.

Violent videos surfaced on X amid Ecuador's 'internal armed conflict'. Taylor Swift's deepfakes prompted calls for criminalising deepfake pornography in the USA. India warned social media companies that they will be held responsible for disseminating AI-generated deepfakes on their platforms.

Development

Germany and Namibia, co-facilitators of the Summit of the Future, announced the release of the zero draft of the Pact for the Future, which includes provisions related to the impact of digital technologies on peace and security, as well as the potential of science, technology, and innovation in advancing sustainable development. Tech companies and governments have pledged to increase their actions to address the climate crisis through the Green Digital Action track at the 28th Conference of the Parties (COP28). Thailand revealed an initiative to enhance human capital and digital development. Iran is grappling with escalating internet costs and censorship despite promised free access. Kyrgyzstan and China signed a digital cooperation agreement to advance technological progress, while the UK and India announced cooperation on digital sustainability. Statistics Netherlands developed a new methodology to map European e-waste. World Bank's recent brief highlighted that enhanced internet accessibility in Nigeria and Tanzania has notably diminished extreme poverty.

THE TALK OF THE TOWN – GENEVA

UNCTAD held its flagship eWeek 2023 on 4 - 8 December 2023. Themed 'Shaping the future of the digital economy', it focused on addressing the oversight of digital platforms and AI, promoting sustainability in the digital economy, enhancing women's digital entrepreneurship, and accelerating the digital readiness of developing countries. UNCTAD collaborated with Diplo in providing the first-ever AI-human hybrid reporting for the event.

Held from 22 January to 2 February 2024, the first cluster of the International Telecommunication Union (ITU) Council Working Group (CWG) and Expert Group (EG) meetings started preparations for the matters to be discussed during the June ITU Council. One highlight is the CWG on Child Online Protection (COP), during which the Secretary-General presented the progress report on the COP Initiative, debriefing on ITU's capacity-building activities and cross-sector partnerships in programmes such as Protection through Online Participation (POP), COP in Sports, and EQUALS Tech4Girls.

12 AI and digital predictions for 2024

2024 will be the year of ***AI***. AI technology will continue to grow deeper through powerful foundational models and wider through more connections to the rest of the digital ecosystem (e.g. IoT, virtual reality, and digital infrastructures). Smaller AI and open-source models will gain traction for their transparency, adaptability, and eco-friendliness.

AI risks will dominate the governance and regulatory debate. Existing risks (e.g. jobs, misinformation, biases) will receive more attention than existential risks. AI governance will become more specific and concrete, addressing computational power, data and knowledge, algorithms, and AI applications.

The push for national sovereignty over data, AI, and technology infrastructure will reshape ***digital geopolitics***. The digital decoupling between China and the USA will accelerate significantly in 2024. India, Brazil, South Africa, Singapore, Türkiye, and Gulf states, among others, will try to carve an asymmetric 'third digital space' between the two superpowers.

There will be a push for new organisations, commissions, and expert groups dealing with AI and ***digital governance.*** The adoption of the UN Cybercrime Convention is anticipated to mark the beginning of global digital governance in 2024. The Global Digital Compact (GDC) will be negotiated, and international bodies, including the UN Security Council, will address digital aspects of conflicts and humanitarian crises.

AI will put ***diplomacy*** to the test through the automation of tasks using Large Language Models (LLMs). Diplomacy will need to adapt to increased pressure to negotiate AI and digital topics, extending to areas such as digital governance, health, trade, and human rights.

Cybersecurity will be a significant concern, with a focus on military conflicts in Gaza and Ukraine, emerging threats to digital critical infrastructure, and AI-facilitated theft and illegal activities. The expected adoption of the UN Cybercrime Convention in early 2024 will further shape global cybersecurity efforts.

AI's impact on ***human rights***, including freedom of expression and privacy, will be on the centre stage, sparking debates on dignity and humanity's definition. Neurorights, influenced by AI and biotechnological developments, will gain prominence.

AI will accelerate ***economic changes***, from restructuring traditional industries to developing new ones built around AI technology. The primary policy dynamics will be related to the economic consequences of the digital decoupling (or de-risking) between China and the USA, anti-monopolies in AI, taxation of online industries, and digital trade.

The standardisation community will prioritise ***standards*** for AI, 6G networks, quantum computing, and other advanced technologies. Multilateral processes will explore technical standards as a soft regulation approach.

The decades-long saga around online ***encryption*** will gain new momentum with the debate around on-device scanning of communications for illegal content.

Digital identities will gain importance, with a focus on the Digital Public Infrastructure (DPI) initiative endorsed by the G20.

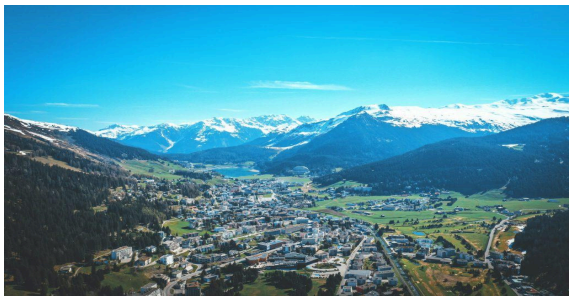
Elections in over 70 countries will heighten concerns about ***online content***, specifically mis/disinformation through deepfake videos, texts, and sounds, which will accelerate efforts to detect AI-generated content.

AI will trigger innovation in new aspects of ***inclusion*** as countries and communities will aim to develop bottom-up AI solutions that reflect their cultural and knowledge heritage.

Read the 12 AI and Digital Predictions for 2024 in full.

AI optimism in geopolitically pessimistic Davos

In the serene backdrop of the Swiss Alps, the World Economic Forum (WEF) in Davos stands as a barometer for the year's technological zeitgeist. After an exceptionally tech-shy Summit in 2023, AI restored tech optimism in Davos this year. The blueness of the tech sky stands out even more against the gloomy and cloudy global geopolitics. Is this optimistic dose of AI a panacea for global problems, the maturation of AI discourse, or yet another tech sleepwalk?



Optimistic dosage. The 24 AI-centric sessions out of 235 at WEF featured 135 speakers and a flurry of 1,101 arguments, with the majority having a positive tone (600). The WEF AI debates have been rated a solid *8* on DiploAI's optimism scale. Conversations have revolved around AI's potential to elevate productivity, combat diseases, and solve environmental crises. The economic forecasts present AI as a trillion-dollar boon, a narrative contrast to AI doomsday scenarios from the spring of 2023.

From extinction to existing risks. The narrative surrounding AI risks has shifted at WEF, with existential risks now viewed as problems humanity can overcome, similar to past technological challenges. Most of the AI risk debate focused on the existing risks of AI misuse, such as misinformation, job losses, and educational reform.

It's unclear why AI gurus have recalibrated their language at WEF, as their predictions that AI would destroy humanity were made with great conviction last year. Sam Altman's frank admission, 'no one knows what comes next [with AI]', encapsulates the dual nature of this juncture: it's both alarming and reassuring. The uncertainty voiced by those at the forefront of AI development is concerning. At the same time, it is encouraging that tech leaders speak more frankly about their knowledge of the impact of AI without the fear-mongering of the last year.

IPR and content for AI. The *New York Times*' court case against OpenAI over the use of copyrighted material to train AI models was frequently mentioned in Davos. Traceability and transparency in AI development will be critical for a sustainable and functional AI economy.

AI governance. Last year's narrative that governance should focus on AI capabilities has given way to a focus on applications. It makes AI less unique and more governable, just like any other technology. This approach, used by the EU AI Act, is also gaining popularity in the USA. The WEF discussions revealed more similarities than differences in the regulation of AI in China, the USA, and the EU.

Open source AI. The stance on open-source AI as an unmanageable risk softened at WEF. Yann LeCun of Meta argued that open-source AI is beneficial not only for scientific progress but also for controlling the monopolies of large AI tech companies and incorporating diverse cultural and societal inputs into AI development. Open-source AI will gain traction in 2024, posing a significant challenge to proprietary software such as OpenAI.

AI and development. According to the UN Secretary-General's Envoy on Technology, Amandeep Sing Gill, AI will not save the SDGs if current trends continue. This candid assessment was more of an outlier in WEF discussions. For example, there was little discussion of the AI-driven widening of digital divides and the increasing concentration of economic and knowledge power in the hands of a few companies.

The confusion around AI potentials and risks underscores the need for a mature, nuanced conversation on AI's future. We must navigate the known challenges with agile, transparent, and inclusive regulatory frameworks. The Davos debates have made strides in this direction, aiming to steer us away from a dystopian AI future through informed, balanced dialogue rather than fear.

Browse session reports from WEF's 2024 annual meeting and the final report from the event.

The intellectual property saga: AI's impact on trade secrets and trademarks

Within the realm of AI and intellectual property, trade secrets and trademarks present unique challenges and opportunities that require special attention in the evolving legal landscape. This blog outlines the complexities, challenges, and opportunities that define intellectual property rights and obligations, focusing on examples from the EU and US legal frameworks.

US trade secret law safeguards diverse information, such as financial, business, scientific, and technical data, if the owner takes reasonable security measures to maintain secrecy. The information must derive value from not being widely known or easily accessible by others through legitimate means (18 U.S.C. §1839(3)). Requirements related to secrecy exclude trade secret protection for AI-generated outputs that are not confidential, such as those produced by systems like ChatGPT or Dall·E. Nevertheless, trade secret laws seem to be more flexible to safeguard various AI-related assets. Namely, there is no stipulation that a trade secret must be originated by a human being, while AI-generated material is treated like any other form of information, as evident in 18 U.S.C. §1839(4), which defines trade secret ownership.

AI innovators often choose trade secret protections over patents due to ambiguous AI and copyright laws. From the EU perspective, the impending AI Act might require disclosure of AI operations, impacting the viability of trade secret safeguarding in some cases. Clear guidelines for AI trade secrets and defining collaboration obligations are crucial for fostering innovation and protecting business assets.

AI integration transforms trademark protection, expanding from logos to AI-generated content and algorithms. Defining the 'average consumer' and determining responsibility in trademark infringement cases pose challenges in AI-driven customer service. There aren't any known cases addressing AI and liability in trademark infringement. However, the Court of Justice of the European Union (CJEU) ruled in cases like Louis Vuitton vs Google France and L'Oréal vs eBay. These decisions state that Google and eBay cannot be held accountable for trademark infringement unless they are aware and actively involved in the automatic selection that results in trademark infringement. Therefore, if we were to apply these cases in AI systems, AI providers would be held liable in the EU if the AI provider plays a more active role in potential infringing actions.

The impact of AI on industries highlights the need for flexible intellectual property laws to balance innovation and protection. Legal frameworks must adapt to recognise and safeguard AI-generated innovations, raising questions about authorship and ownership attribution. Policymakers and stakeholders must craft forward-thinking regulations to accommodate AI's potential while protecting all parties' rights and interests.

A longer version of this blog first appeared on the Digital Watch Observatory. Read the full version of the blog, and discover part 1 of the blog series, The intellectual property saga: The age of AI-generated content.

In focus

The perfect cryptostorm

We watched the new Netflix documentary 'Bitconned' on cryptocurrency industry fraud. Read what were the ingredients to create such a perfect storm for victims.

The cryptocurrency and blockchain craze from 2017 to 2021 unfolded in a unique setting. One component amplified the other, multiplying the effect, thus creating a perfect cryptostorm that impacted trust in the industry and caused financial losses.

Bitcoin, often termed the digital gold, stood out as a marvel of human engineering and a one-hit wonder. This led to the emergence of a new payment industry, driven by legacy financial organisations seeking relevance in the digital era.

Simultaneously, the retail investing industry saw an influx of capital, with online trading companies backed by institutional investors posing risks for retail users and consumer protection. Unanswered risks, changes in the financial industry, and inexperienced investors set the stage for a perfect storm, exacerbated by human greed.

The perfect cryptostorm. The Netflix documentary 'Bitconned' vividly portrays the summoning of the cryptostorm by companies like Centra Tech. In 2017, Centra Tech raised $25 million for a VISA-backed credit card, only to be revealed as a staged mirage. The court case concluded in 2021, resulting in jail sentences. The documentary, led by Ray Trapani, one of Centra Tech's key figures, reveals how young scammers raised millions in an ICO with a one-page website.

The cryptostorm persisted for years, culminating in the collapse of FTX, the world's second-largest cryptocurrency exchange. Companies like Celsius and Luna also faced legal challenges for misusing investor funds.

How did crypto scam companies utilise the above ingredients? By promising the right thing at the right moment. Internet users witnessed the financial sector's transformation and bitcoin's success. They could easily be convinced that a new decentralised finance infrastructure is on the verge, which will be supported by the lack of a regulatory framework. At the same time, giving them a fair chance to participate in the industry beginnings and become the new crypto millionaires, which was the main incentive for many. If people behind the open-source cryptocurrency (bitcoin) could create the 'internet of information', the next generation of cryptocurrency engineers would surely deliver the 'internet of money'. However, again, it was false. It was, in fact, a carefully worded money-grabbing experiment.

All the above ideas still stand as a goalpost for further industry developments.

Could this happen again for online financial services? Looking ahead, the likelihood of a recurrence of such large-scale scams within online financial services is slim. Regulatory frameworks have been strengthened, and authorities are better equipped to identify and intervene in fraudulent activities. Moreover, investors have become more discerning, and wary of promises that seem too good to be true.

However, the potential for deception remains inherent in any technology-driven industry, as opportunistic actors may continue to exploit societal aspirations for innovation and progress. As such, vigilance and scepticism are essential when engaging with new technologies and investment opportunities, ensuring that users are not misled by false promises or exaggerated claims.

A longer version of this blog first appeared on the Digital Watch Observatory. Read the full version of the blog.

OEWG wraps up its sixth substantive session

The sixth substantive session of the UN Open-Ended Working Group (OEWG) on security of and the use of information and communications technologies 2021–2025 was held in December 2023, marking the midway point of the process. Here is a quick snapshot of the discussions.

Threats. The risks and challenges associated with emerging technologies, such as AI, quantum computing, and the internet of things (IoT), were highlighted by several countries. Numerous nations expressed concerns about the increasing frequency and impact of ransomware attacks on various entities, including critical infrastructure, local governments, health institutions, and democratic institutions. Many countries emphasised the importance of international cooperation and information sharing to effectively address cybersecurity challenges. The idea of a global repository of cyber threats, as advanced by Kenya, enjoys much support.

Rules, norms and principles. Delegations signalled that clarifying the norms and providing implementation guidance is necessary. Delegations expressed different views on whether new norms are needed.

International law. The elephant in the room is the question of whether a new treaty and new binding norms are needed. The law of state responsibility, the principle of due diligence, and the applicability of international humanitarian law (IHL) and international human rights law in cyberspace are also areas without consensus.

Confidence-building measures (CBMs). There's widespread support for the global Points of Contact (PoC) directory as a valuable CBM. The OEWG members will focus on its implementation and operationalisation.

Capacity building. Foundational capacities such as legal frameworks, the establishment of dedicated agencies, and mechanisms for incident response, with a special focus on computer emergency response teams (CERTs) and CERT cooperation, were consistently highlighted as crucial. Delegations also stressed the importance of national contexts and how there is no one-size-fits-all answer to building foundational capacities. Delegations expressed support for the voluntary cybersecurity capacity-building checklist proposed by Singapore and for the Accra Call for Cyber Resilience Development set forth during the Global Conference on Cyber Capacity Building (GC3B).

Regular institutional dialogue. The discussions on what the future regular institutional dialogue will look like can be summarised as Programme of Action (PoA) vs OEWG. The supporters of the Programme of Action (PoA) suggest using the review mechanism to identify gaps in existing international law and recognise that such gaps can be filled with new norms. Those who support continuing regular institutional dialogue within the framework of OEWG suggest a permanent OEWG should be established, focusing on developing legally binding rules as elements of a future universal treaty on information security. They propose consensus-based decision-making and stricter rules for stakeholder participation.

About this issue: Issue 86 of the Digital Watch monthly newsletter, published on 5 February 2024 by the Geneva Internet Platform and DiploFoundation. For more coverage and analysis, visit the Digital Watch observatory
Team: Andrijana Gavrilović (author), Boris Begović (contributor), Arvin Kamber (contributor), Bojana Kovač (contributor), Yung-Hsuan Wu (contributor), Ginger Paque (editor), Diplo's CreativeLab (design) | Get in touch: digitalwatch@diplomacy.edu
On the cover: *AI gurus: between pyromaniacs and firefighters.* Credit: Vladimir Veljasević DiploFoundation (2024)

The Geneva Internet Platform is an initiative of:

GENEVA INTERNET PLATFORM digwatch
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
REPUBLIQUE ET CANTON DE GENEVE
DIPLO www.diplomacy.edu