EU Digital Technologies and Policy Conference (EUDTP 2025) Abstracts and Contributions

Juan-Antonio Cordero-Fuertes, Mehwish Alam, Olivier Blazy, Anne Alombert, Natalia Díaz-Rodríguez, Teodora Curelariu, Pratiksha Ashok, Calina Ciuhu, Stefano de Luca, Claudio Feijóo, et al.

▶ **To cite this version:**

Juan-Antonio Cordero-Fuertes, Mehwish Alam, Olivier Blazy, Anne Alombert, Natalia Díaz-Rodríguez, et al.. EU Digital Technologies and Policy Conference (EUDTP 2025) Abstracts and Contributions. EU Digital Technologies and Policy (EUDTP) Conference 2025, May 2025, Bruxelles, Belgium. 2025. ⟨hal-05267621⟩

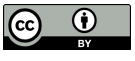
HAL Id: hal-05267621

https://hal.science/hal-05267621v1

Submitted on 28 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

EU Digital Technologies and Policy Conference

EUDTP 2025

Abstracts and Contributions

Juan-Antonio Cordero-Fuertes (Ed.)

Contents

Foreword

It was my distinct pleasure to open the second edition of the EU Digital Technologies and Policy Conference (EUDTP) last spring in Brussels. I was especially proud to represent *Institut Polytechnique de Paris*, the official organizing institution of this event, and to welcome such an outstanding group of researchers, professionals, and policymakers from across Europe.

This year's edition brought together over 100 participants, with 30 speakers from more than 20 institutions contributing across six sessions. The conference offered a rich and interdisciplinary program-ranging from advances in AI and digital infrastructures to energy, sustainability, regulation, and the societal impact of emerging technologies. I was particularly delighted by the strong turnout, the quality of the presentations, and the insightful discussions that followed.

The success of this edition reaffirms the growing importance of EUDTP as a platform for open, cross-sector collaboration-bringing together academia, industry, EU institutions, and the broader public to engage with Europe's digital challenges and opportunities. The active involvement of the European Parliament and its Research Service (EPRS), along with EuroTech, EuroTeQ, and IEEE, reflects the value of building strong and lasting partnerships between science and policy.

At *Institut Polytechnique de Paris*, we are committed to exploring the possibility to develop EUDTP into a long-term, interdisciplinary European community for exchange and collaboration-spanning domains such as AI, cybersecurity, quantum computing, and the enabling infrastructures and resources that support them.

I would like to warmly thank all the experts who contributed their time, insight, and expertise, and the entire organizing team –especially our colleagues at IP Paris– for delivering such a well-run and impactful event. We look forward to building on this momentum in the years ahead.

Christopher Cripps
Vice-President for Europe and International Affairs
Institut Polytechnique de Paris

Editor's Note

The EU Digital Technologies and Policy conference (EUDTP 2025) took place in Brussels on May 14th and May 15th. The conference was organized by the *Institut Polytechnique de Paris*, with the support of the EuroTech Universities Alliance and the EuroTeQ Engineering University, the European Parliamentary Research Service (EPRS), and the technical co-sponsorship of the IEEE Society of Social Implications of Technology (SSIT). I want to thank all involved institutions for their commitment and support, as well as the members of the conference committees (listed below) for their work, advice and energy dedicated to EUDTP 2025.

This document includes contributions of speakers invited to the conference, and contributions submitted to the public call for abstracts of the conference. Submitted abstracts (18 in total) were peer-reviewed by the program committee; 7 of them were selected to be presented in EUDTP 2025, and 5 more abstracts could not be presented in the conference (due to time limitations), but given their interest, were accepted for inclusion in this proceedings document. The nature of each contribution (*invited contributions* and *regular contributions* presented in the conference, and *abstracts/contributions accepted for publication*, not presented in the conference) is precised on the corresponding section. Contact details of speakers and authors are also included.

Juan-Antonio Cordero-Fuertes
EUDTP 2025 General Chair
École Polytechnique – Institut Polytechnique de Paris

Steering and Program Committees

Steering Committee Sonia Vanier, École Polytechnique – IP Paris, France.– Hervé Debar, Télécom Sud Paris – IP Paris, France.– María Montoiro, Institut Polytechnique de Paris (IP Paris), France.– Jay Pearlman, IEEE Society on Social Implications of Technology (SSIT), United States.– Anita Schneider, EuroTech Universities Alliance, EU (Belgium).

Program Committee Ruta Binkyte, CISPA, Germany.– Cristina Blasi, Universidad Autónoma de Barcelona (UAB), Spain.– Gregory Blanc, Télécom Sud Paris – IP

Paris, France.– Miguel Colom, ENS Paris-Saclay, France.– Marceau Coupechoux, Télécom Paris, France.– Claudia D'Ambrosio, CNRS, France.– Jordi Domingo-Pascual, Universidad Politécnica de Cataluña (UPC), Spain.– Margarita González, Georgia Tech Research Institute (GTRI) and IEEE, United States.– Juan Herrera, Universidad Politécnica de Madrid (UAM), Spain.– Darina Martikanová, Universidad Autónoma de Madrid (UAM), Spain.– Eduardo Oliva, Universidad Politécnica de Madrid (UPM), Spain.– Alonso Silva, Nokia Bell Labs, France.

Contributors and Speakers

Speakers and main authors of contributions included in this document Mehwish Alam, Télécom Paris – IP Paris, France.– Anne Alombert, Université Paris 8, France.– Pratiksha Ashok, Tilburg University, The Netherlands.– Olivier Blazy, École Polytechnique – IP Paris, France.– Calina Ciuhu, TU Eindhoven, The Netherlands.– Teodora Curelariu, Université Grenoble Alpes & INRIA, France.– Stefano De Luca, EPRS, EU (Belgium).– Natalia Díaz-Rodríguez, Universidad de Granada, Spain.– Claudio Feijóo, Universidad Politécnica de Madrid (UPM), Spain.– Neringa Gaubienė, Vilnius University, Lithuania.– Bissan Ghaddar, DTU, Denmark & Ivey Business School at Western University, Canada.– Fabio Giglietto, Università degli Studi di Urbino Carlo Bò, Italy.– Rafael Gomà, Barcelona Supercomputing Center (BSC), Spain.– Gloria González-Fuster, VUB, Belgium.– Arturas Grumulaitis, Vilnius University, Lithuania.– Petia Guintchev, Universidad de Barcelona (UB), Spain.– Romain Jacob, ETH Zurich, Switzerland.– Laima Jančiūtė, Universiteit van Amsterdam (UvA), The Netherlands.– Vicky Kalogeiton, École Polytechnique – IP Paris, France.– Georges Kariniotakis, Mines Paris – PSL, France.– Juan Knaster, Fusion for Energy, EU (France).– Luise Koch, TU Munich, Germany.– Philipp Kreer, TU Munich, Germany.– Kim Krüger, TU Munich, Germany.– Diane Leblanc-Albarel, KU Leuven, Belgium.– Jukka Manner, Aalto University, Finland.– Andrew McStay, Bangor University, United Kingdom.– María Ortiz de Zúñiga, Fusion for Energy, EU (Spain).– Patrice Nivaggioli, Cisco Systems France, France.– Ivanka Popović, University of Belgrade, Serbia.– Simona Ramos, UPF Barcelona School of Management (BSM), Spain.– Markus Roth, Focused Energy GmbH & TU Darmstadt, Germany.– Jochen Spangenberg, Deutsche Welle (DW), Germany.–

Chapter I

AI Technologies, Services, and Applications

I.1 Connecting the Dots in Trustworthy Artificial Intelligence

Invited contribution
Natalia Díaz-Rodríguez, nataliadiaz@ugr.es*, Universidad de Granada, Spain*

Trustworthy technical requirements are based on three main pillars that should be met throughout the system's entire life cycle: it should be (1) lawful, (2) ethical, and (3) robust, both from a technical and a social perspective.

However, attaining truly trustworthy AI concerns a wider vision that comprises the trustworthiness of all processes and actors that are part of the system's life cycle, and considers previous aspects from different lenses. A more holistic vision contemplates four essential axes: the global principles for ethical use and development of AI-based systems, a philosophical take on AI ethics, a risk-based approach to AI regulation, and the mentioned pillars and requirements.

The seven requirements (human agency and oversight, robustness and safety, privacy and wellbeing, and accountability) are analyzed from a triple perspective:

- *What* each requirement for trustworthy AI is,
- *Why* is it needed, and
- *How* each requirement can be implemented in practice.

On the other hand, a practical approach to implement trustworthy AI systems allows defining the concept of responsibility of AI-based systems facing the law, through a given auditing process. Therefore, a responsible AI system is the resulting notion we introduce in this work, and a concept of utmost necessity

that can be realized through auditing processes, subject to the challenges posed by the use of regulatory sandboxes. Our multidisciplinary vision of trustworthy AI culminates in a debate on the diverging views published lately about the future of AI. Our reflections in this matter conclude that regulation is a key for reaching a consensus among these views, and that trustworthy and responsible AI systems will be crucial for the present and future of our society.

I.2 LLMs and Symbolic AI

Invited contribution
Mehwish Alam, `mehwish.alam@telecom-paris.fr`*, Télécom Paris – IP Paris, France*

Large Language Models (LLMs) have significantly transformed the landscape of AI, being extensively employed in various Natural Language Processing (NLP) tasks, including natural language understanding, question answering, and machine translation. However, LLMs are susceptible to hallucinations, where the generated output may appear factual but lacks grounding, posing challenges to their applications to real-world scenarios. Recent efforts within the NLP community have started addressing this issue by introducing techniques for automatically detecting and mitigating hallucinations as well as explainability across various NLP tasks. This talk gives an overview on various aspects and challenges which are currently being addressed in this field along with the role of Symbolic AI in addressing those challenges.

I.3 AI Conditional Content Generation for Multimodal Data

Invited contribution
Vicky Kalogeiton, `vicky.kalogeiton@polytechnique.edu`*, École Polytechnique – IP Paris, France*

In this talk, I present fundamental research in cross-modal conditional generation with structured outputs, spanning multiple modalities and architectures. I briefly present a series of recent breakthroughs in traditional multimodal data, such as image and video generation from text, video question-answering. Then, I present multimodal works applied to spatial domains, for instance for camera placement and visual geolocalization.

I.4 Unboundling Social Networks and Collaborative Recommendation

Invited contribution
Anne Alombert, `anne.alombert@univ-paris8.fr`*, Université Paris 8, France*

In this talk, I provide a brief overview of the challenges of disinformation in the context of commercial social networks and generative artificial intelligence, and I suggest two levers to overcome the current situation, in terms of regulation and innovation.

I first argue that our current disinformation problem is linked to the business model of digital platforms, based on data economy and attention economy, as well as to their technological functionalities, based on automatic content recommendations as well as personalized profiling and smart targeting. The rapid development of generative artificial intelligence applications could definitively worsen the situation, by allowing a massive generation of fake contents as well as the constant feeding of fake accounts and bots. In such a context, not only are electoral processes disrupted (as we have recently seen in the United States with X or in Romania with TikTok and Facebook), but it will also become increasingly complicated for citizens to trust the digital contents circulating in digital media, whereas these same digital media now constitute the main sources of information, especially for younger generations. As a result, the very possibility of public debate and democratic life are threatened.

To remedy this problem, I suggest two levers, in terms of regulation on the one hand, and in terms of innovation on the other. The first lever is the unbundling or degrouping of social networks, which would allow us to remove the hegemonic power on content recommendation from digital giants and open space for new stakeholders, which would make possible to reach an "algorithmic pluralism" in the digital sphere. The second lever is collaborative recommendation, which would allow citizens to regain control over the configuration of their daily information environments, through the collective and democratic selection of the contents they want to share.

These two levers are both promoted by the French Digital Council and by many stakeholders form civil society. Faced with the alliance between libertarian tech companies and ultra-liberal authoritarian regimes that we are witnessing today, these two perspectives would benefit from being supported at the European level, in order to ensure the possibility of democratic and independent debate in the European space.

I.5 AI to Accelerate Fusion Engineering

Invited contribution

María Ortiz de Zúñiga, `Maria.Ortiz-De-Zuniga@f4e.europa.eu`*, Fusion for Energy, EU (Spain)*

For many years, Artificial Intelligence (AI) has been applied to many different fields, such as medicine, robotics, linguistics, data mining, decision-making, videogames and the automotive industry, whereas in others it has recently started being explored. In general, this is the case of the nuclear manufacturing and, specifically, the case related to quality control in manufacturing of

the large ITER Vacuum Vessel (VV). Fusion for Energy (F4E) has been applied to accelerate the validation and analysis of outputs from the phased-array ultrasonic (PAUT) non-destructive testing (NDT) and to show how to enhance quality control by predicting of the weld success rate through the development and analysis of AI tools applied to the ITER Vacuum Vessel manufacturing. Other practical applications include prediction of material properties for fusion-specific materials, such as CuCrZr, and other reverse engineering applications in superconducting magnets design. All these practical AI developments have allowed fusion engineers to find trends where statistics could not, to predict the success rate of first-of-a-kind operations and to qualify new materials to withstand the harsh environment where fusion reactions take place. Collaboration with nuclear code designers and engineers, showing the success of these developments, is enabling the discussion to include these new technologies in codes and standards.

Chapter II

Digital Infrastructures

II.1 Submarine Communication Cables: Security and Resilience in the EU

Invited contribution
Stefano De Luca, `stefano.deluca@europarl.europa.eu`*, EPRS, EU (Belgium)*

The submarine cable network, comprising fibre optic cables laid on the ocean floor, digitally connects countries worldwide. The European Union (EU) recognizes the strategic importance of submarine cable infrastructures for its digitalization and economy, as over 99% of international data traffic rely on them. However, these infrastructures face two primary challenges: physical attacks (e.g. cutting submarine cables) and cyber threats, including foreign technology dependency (e.g. which can lead to data interception, surveillance, and internet traffic disruption by foreign-state actors).

Repairing damaged submarine cables is a complex task due to two main reasons. First, the damage may be located in deep waters or under ice, making it difficult to access. Second, the availability of cable repair ships is limited, and their on-demand deployment is not guaranteed. On the cyber threats/technology dependency front, various reports have accused foreign actors of tapping into submarine cable networks to spy on other countries.

In response to these concerns, the Commission published a recommendation in February 2024 on the security and resilience of submarine cable infrastructures. The key proposed actions include:

(i) Establishing a Cable Infrastructure Expert Group, comprising Member States' authorities, to serve as a coordination platform for information exchange, mapping existing submarine cables, and assessing EU-wide risks and vulnerabilities.

(ii) Creating a Cable Security Toolbox to recommend measures to reduce risks, based on a comprehensive EU-wide assessment of the sector, including

strategies for dealing with high-risk suppliers.

(iii) Developing a list of projects for the deployment of strategic submarine cables that meet specific criteria, such as mitigating dependencies on high-risk entities, with financing options including EU investments and funding.

Notably, 70% of submarine cable incidents are unintentional and could be reduced by establishing cable protection zones that restrict certain activities by commercial vessels (e.g. anchoring and fishing) near the cables. Nevertheless, physical sabotage and cyberattacks by foreign actors remain significant concerns. To address these concerns, various authors have suggested a different set of actions such as:

1. Utilizing the European Defence Fund to invest in submarine cables' protection technologies, such as sensors and detection systems.
2. Requiring Member States to incorporate the use of detection systems into licence requirements for landing submarine cables.
3. Allocating a dedicated budget envelope to enable governments to invest in protecting their submarine cable infrastructure.
4. Investing in submarine cable repair capabilities by funding an EU equivalent of the U.S. Cable Security Fleet.
5. Developing a comprehensive and common approach to support EU-based companies in constructing new secure submarine cable routes.

II.2 Understanding the Numbers in Sustainability Discussions

Invited contribution
Romain Jacob, `jacobr@ethz.ch`*, ETH Zurich, Switzerland*

It is encouraging to see more and more studies published about the environmental footprint of the ICT sector. Unfortunately, the outcomes of those studies are often misinterpreted. In fact, one can look at the footprint of a product or activity in many different ways which all make sense but serve different purposes. It is very easy to mistake one purpose for another and thus derive completely wrong conclusions, which may lead to harmful-albeit well-intentioned-decision-making.

The best way to avoid those misunderstandings is to clarify the different approaches and their corresponding purpose.

This mental framework helps draw correct conclusions from the growing corpus of sustainability studies.

In this talk, I present three of the most important methodological choices. I then conclude with a couple of examples from computer networks (my area of research) to illustrate how easy it is to misinterpret footprint numbers.

II.3 ChatControl EU Regulation: Impact and Risk

Regular contribution
Diane Leblanc-Albarel (speaker, `diane.leblanc-albarel@kuleuven.be`*), joint work with Bart Preneel, KU Leuven, Belgium*

This talk presents the proposed European regulation known as "ChatControl", which aims to combat the dissemination of Child Sexual Abuse Material (CSAM) through mandatory scanning of private communications.

Child Sexual Abuse Material (CSAM) typically consists of images or videos depicting the abuse of minors. CSAM are divided in two categories: (1) known material already identified but still circulating online, and (2) newly created content. In recent years, the US' National Center for Missing and Exploited Children (NCMEC), a leading organization coordinating efforts against CSAM, has reported an exponential increase in online CSAM reports. This surge underscores the urgency of the problem and the need for measures to address it.

The ChatControl regulation, discussed at the European level for three years, proposes mandatory scanning of communications by service providers (e.g., WhatsApp, Signal, Telegram, Gmail) to report suspected CSAM sharing. By making the detection and reporting of CSAM a legal obligation for service providers, the regulation aims to significantly reduce the proliferation of this material and strengthen protections for children.

If the ChatControl regulation were adopted in its current form, compliance by service providers would necessitate implementing either backdoors in encrypted communications or client-side scanning on users' devices. The first option, which would compromise the integrity of end-to-end encryption (used by services like WhatsApp and Signal), has been widely criticized as infeasible due to the privacy issues it poses. As a result, client-side scanning has emerged as the more viable alternative. This approach involves scanning content directly on users' devices before it is sent.

Today, service providers can voluntarily detect and report CSAM. To do so, they use so-called perceptual hash functions. These functions identify known CSAM by generating unique "fingerprints" of content and comparing them against databases of flagged CSAM material. Perceptual hash functions can detect slightly altered content, making them effective tools for detecting content that has been slightly altered but still represents the same material. If the regulation were accepted, these functions would be employed on a massive scale.

This talk presents the current capabilities of perceptual hash functions, their strengths, and their limitations. While these tools have demonstrated effectiveness in detecting known CSAM, their weaknesses –particularly in terms of false positives– pose serious risks. A high rate of false positives could result in individuals being wrongly accused of possessing CSAM. The presentation provides published quantitative estimates of these risks. The technical challenges discussed will be supported by several scientific papers, particularly our recent work on the subject [1].

Beyond the technical challenges and privacy concerns, the ChatControl reg-

ulation presents profound societal dilemmas. The automatic scanning of European citizens' devices risks setting a precedent for deeper intrusions into private lives. Even if advancements reduce false positives, this regulation presents a crucial dilemma. This talk aims to spark a debate on the societal values at stake: not just how we should balance public safety with individual privacy and fundamental rights, but where do we, as a society, want to draw the line.

References

1 Diane Leblanc-Albarel and Bart Preneel, "Black-box Collision Attacks on the NeuralHash Perceptual Hash Function", Cryptology ePrint Archive. Available at: https://eprint.iacr.org/2024/1869

II.4 Datacenter Energy Footprint and LLM Power Inference

Invited contribution
Patrice Nivaggioli, pnivaggi@cisco.com*, Cisco Systems France, France*

Data centers, which host LLM inference, contribute significantly to global energy consumption and carbon emissions. Optimizing energy efficiency is essential to reduce the carbon footprint of AI technologies as well as lower cost, making LLMs more accessible for users. After reviewing LLM inference phases and characteristics, we present various optimisation strategies concerning energy efficiency and including software/hardware optimisation, workload management, model architectures and inference frameworks. Finally we provide further research directions and recommendations for future explorations in the field of sustainable LLM inference.

II.5 Life-Cycle Emissions of ICT Infrastructures

Invited contribution
Jukka Manner, jukka.manner@aalto.fi*, Aalto University, Finland*

The ICT industry and ICT services have an environmental footprint from the energy usage and the manufacturing of the devices and hardware. The increased energy usage is becoming a major concern due to the increase in emissions from the production of electrical energy. Yet, as we move toward low carbon renewable energy, this impact will hopefully diminish in the long term. Thus, in the future, the main environmental footprint of ICT is related to hardware manufacturing and the lifetime of the devices.

The ICT hardware is composed of tens of different materials and minerals, and some are in so small quantities in the embedded electronics that recycling is very difficult or even impossible with current knowhow. The repairability

of the hardware is often very limited and sometimes impossible, which means that keeping hardware running becomes a major challenge for both consumers and industry. Planned obsolescence, availability of spare parts, difficult and expensive repairability and limitations in software updates further aggravates the problems.

This talk advocates more work and regulation towards extending the lifetime of ICT hardware, for the benefit of all the industries and public institutions using ICT services, and consumers.

Chapter III

Materials and Physical Basis of Computing

III.1 RISC-V: the Opportunity for Europe to Recover Chips Sovereignty

Invited contribution
Rafael Gomà, `rafael.gomatorrellas@bsc.es`*, Barcelona Supercomputing Center (BSC), Spain*

The RISC-V open standard is becoming one of the key levers to reduce external dependencies of foreign semiconductor IP and support European innovation and industry. In this talk we revisit the concept of RISC-V and why it is important for Europe. We review where Europe stands on the adoption of this technology and point out some success stories. We discuss some of the positive aspects of the European initiatives and what aspects should be further addressed to reduce the gap on chip design and manufacturing capabilities. Finally, we provide a quick overview of the BSC projects involving RISC-V.

III.2 Navigating Complexity in High Precision Semiconductor Systems

Invited contribution
Calina Ciuhu, `c.ciuhu@tue.nl`*, TU Eindhoven, The Netherlands*

The semiconductor industry is continuously driven by the need for higher throughput, precision mechatronics, sustainable processes, and lifetime monitoring. These demands pose several challenges:

- The pursuit of high precision positioning increases systems complexity. For

instance, introducing smart materials, such as piezoelectric, significantly enhances performance, but introduce non-linearities.

- These systems operate at the intersection of multiple physical domains – mechanical, thermal, and electromagnetic– further complicating modelling efforts and requiring higher computational resources.
- With the rapid advancing developments in artificial intelligence (AI), it becomes interesting to explore new paradigms, however transferring large models assumes large amounts of data.

With increasing requirements towards higher performance indicators, this inevitably implies that our impressions on the system have to be more accurately represented. Collecting more data which does not encompass system's complexity might lead to redundancy. Therefore, the complexity is a complementary part of collecting the evidence. One approach is to regard the system through information flow, or information balance. An intriguing relationship between model evidence (marginal likelihood), complexity, and accuracy can be derived from Bayes' theorem, (evidence) = (accuracy) - (complexity), see Eq. (III.1):

$$\log p(y|m) = \int p(\phi|y,m) \log p(y|\phi,m) d\phi - \int p(\phi|y,m) \log \frac{p(\phi|y,m)}{p(\phi,m)} d\phi \quad \text{(III.1)}$$

This relationship underscores the importance of improving model accuracy. This presentation elaborates on common challenges observed in three relevant use-cases characteristics for the semiconductor industry,

1. Leveraging physics-informed models to enhance precision control and performance
2. Applying information theory to regularize non-unique, integral problems
3. Combining analytical and numerical methods for effective lifetime prediction and monitoring, addressing drifts, aging materials, typically resulting in unstable or boundary conditions ill-posed problems

These examples illustrate how AI can be harnessed to navigate the complexities of modern semiconductor systems, ensuring both high performance and robustness through a balanced information flow, encompassing physics-inspired priors, informative data, and systems complexity.

III.3 The Lithium Mining Project in Jadar and the Green Transition in Europe

Invited contribution

Ivanka Popović (speaker, `ivanka@tmf.bg.ac.rs`*), University of Belgrade, Serbia, joint*

work with Zoran Stevanović, University of Belgrade, Serbia, and Bogdan Šolaja, Serbian Academy of Sciences and Arts (SANU), Serbia

The jadarite mineral deposits found in the Jadar Valley in Serbia are considerable and may, by some estimates, substantially contribute to Europe's current lithium requirements. The planned Jadar lithium mine, a project of the Rio Tinto Corporation, along with its accompanying infrastructure, production facilities and, at least, two landfills is located in the Jadar basin in the western part of Serbia in a highly productive agricultural area. The project would cover a territory populated by about 20,000 people whose livelihood would be threatened. This area is closely linked to the downstream Mačva basin, which represents the most significant groundwater reserve source in Western Serbia. The potential pollution resulting from mining operations, processing technologies, solid and liquid waste disposals and their transportation, could endanger potable water supply in the Jadar area, reaching all the way to the capital city, Belgrade. The planned project is expected to cause significant habitat destruction and fragmentation, resulting in negative impacts on the living world, including several hundred plants and animal species of which 145 are protected or strictly protected. The industrial waste landfill in the Štavica stream basin would lead to the removal of 26 000 m^3 of wood causing the destruction of trees for CO_2 assimilation, soil erosion, the pollution of spring water, the disappearance of wildlife in the basin and riverbed and an increased risk of destructive torrential floods.

The ore processing technology is a highly aggressive one involving the annual digestion of 853,333 t of ore by more than 320,000 t of concentrated sulphuric acid at 80 - 95°C and within a pH range of 2.0-3.8. The resulting digestate would be concentrated and further treated to annually produce 58,000 t of lithium carbonate and 286,000 t of boric acid with 259,000 t of sodium sulphate as byproduct. The waste tailings would contain a high concentration of boron and other hazardous elements such as arsenic and others. Despite the announced innovative technology, the company has been unable to meet legal limits for boron in soil and water. Consequently, these legal limits for boron have been removed for boron in soil, so boron is no longer considered to be a harmful substance according to the newest regulations in Serbia.

The proposed Jadar Valley mine would have an adverse effect on 22 villages that depend on agriculture by devastating their rich agricultural land, some 14,000 apiaries and multiple livestock farms. The Government of Serbia has supported the project from the very beginning and has adapted legislation related to expropriation and land use to facilitate mining projects, including articles related to the expropriation of land and the automatic provision of an exploitation permit after the completion of exploration and the approval of an environmental impact assessment study. However, these changes were met with resistance from environmental activists and the local population and have, for the moment, been temporarily withdrawn by the Government of Serbia.

The potential exploitation of the lithium deposit in Western Serbia, although declared as substantial, is a venture with numerous negative aspects: extensive

environmental risks, endangered water supply and social upheaval. This project is the only one in the world where lithium extraction is planned in a populated and fertile agricultural area and with a high degree of certainty that it will destroy the richest groundwater reservoir in entire Serbia, whose reserves are estimated as almost equal to the actual total national potable water demands. This deposit, representing about 1% of the global lithium reserve, does not offer sufficient amounts that would solve the global climate change problem.

III.4 The Governance of the Post-Quantum Cryptography Transition in the EU

Regular contribution

Laima Jančiūtė (speaker, `l.janciute@uva.nl`*), joint work with Ot Van Daalen and Joris Van Hoboken, Universiteit van Amsterdam (UvA), The Netherlands*

Encryption underpins the security of digital networks and protects data privacy. However, currently used encryption methods (in particular, asymmetric (public key) encryption) are threatened by rapid development of quantum computing. Once cryptographically relevant quantum computers will be available, they are expected to be able to break conventional encryption. To counter this threat, novel methods of encryption are being developed. Post-quantum cryptography (PQC) is considered to be the main method in mitigating the quantum threat to cybersecurity. Initiation of migration to PQC is an urgent issue as information can be intercepted now, stored and retrospectively decrypted later and because the deployment of PQC could take decades.

This talk analysed the governance aspects of the PQC transition in the EU. This process here has been evolving quite differently from the US, for example, where a largely top-down approach has been taken (and has been possible). Although the EU has long been funding the development of PQC solutions and the contribution of European researchers to the international PQC standardisation processes has been significant, a more articulated PQC transition-related policy was formulated at the EU level only in 2024. This presentation explored the relevant provisions in the EU legal acts containing implicit requirements to adopt PQC (e.g. NIS2, CRA, GDPR, DORA, eIDAS2 and other) and how they are complimented by the various policy documents making specific references to PQC, including, but not limited to, the EU Commission recommendation on the stipulation of a coordinated PQC transition roadmap and the on-going work undertaken by the Member States. It explained the complex interplay between national and EU-level competences in which the PQC transition is embedded and discussed the contributions of the various national and EU-level actors in shaping this transition process.

Chapter IV

Energy and Sustainability

IV.1 Enabling the Hydrogen Economy: A Taxonomy of Business Models and the Role of Digital Technologies

Regular contribution
Kim Krüger (speaker, kim.krueger@tum.de*), joint work with Timo Böttcher and Helmut Krcmar, TU Munich, Germany*

The European Union's ambitious climate targets have driven a wave of political directives to reduce emissions across industries. Policies like the EU Emissions Trading Scheme compel businesses to internalise polluting emissions costs, creating more substantial incentives for adopting zero-emission alternatives. Hydrogen plays a critical role in these efforts, with its ability to store renewable energy and decarbonise sectors like heavy industry and transportation. However, the path to a sustainable hydrogen economy is fraught with challenges. Transitioning to green hydrogen requires overcoming high production costs and insufficient technical standards and establishing supportive regulatory frameworks. Economically feasible business models (BMs) are essential to ensure the transformation toward a hydrogen economy remains sustainable in the long term. These BMs must address significant upfront investments, operational costs, and infrastructure requirements while fostering innovation and market competitiveness.

To enable the development of a hydrogen economy that aligns with the EU's decarbonisation goals, this research systematically analyses current and future BMs in the hydrogen economy. A sample of 50 companies was analysed to identify BM design elements and characteristics, which are described in a taxonomy outlining recurring patterns and archetypes of hydrogen BMs. The business model archetypes were validated through ten expert interviews, providing insights into the barriers, challenges, and opportunities within the hydrogen

economy. This comprehensive methodology systematically explores hydrogen BMs' economic, technological, and regulatory dimensions.

The developed taxonomy describes three BM dimensions: value proposition, value creation and delivery, and value capture, and aims to guide BM innovations toward integrating hydrogen. It consists of 16 dimensions and 61 unique characteristics. Based on the classification of 50 BMs, we identify eight archetypal BM patterns: Technology Developer, Advanced Manufacturer, Plug-and-Play Solution Provider, Mobility Solution Provider, Infrastructure Provider, Commodity Producer, Distributor, and Service Orchestrator. Each archetype is characterised by its unique function within the hydrogen economy and is visualised using a value map that provides a clear framework for understanding their roles and relationships. Hydrogen-related BMs are predominantly product-oriented, focusing on technological and operational innovations. However, socio-technical factors, including adopting advanced digital technologies, significantly influence their success.

Analysing the critical economic, technological, and socio-political factors affecting hydrogen adoption highlights how digital technologies and regulations support sustainable and scalable business ecosystems. Digitalisation needs to play a more significant role in current hydrogen BMs. Substantial changes will be required to achieve effective network exchange between individual actors across sectors, as digitalisation can facilitate the implementation of BMs by offering tools such as data-driven optimisation, predictive analytics, and blockchain for secure energy transactions. Furthermore, collaborative approaches that account for the interdependencies among market actors are needed, proposing that integrated solutions are vital for unlocking the full potential of hydrogen within the EU's energy transition framework.

By linking the technological potential of digitalisation with evolving regulatory frameworks, this research provides a roadmap for policymakers, industry leaders, and researchers to foster a resilient hydrogen economy. It underscores the importance of coordinated efforts to ensure innovative and digitally supported BMs accelerate the EU's transition to a sustainable, low-carbon future while driving economic growth and energy security.

IV.2 Magnetic Confinement Fusion in Europe: Present and Perspectives

Invited contribution
Juan Knaster, `Juan.Knaster@f4e.europa.eu`*, Fusion for Energy, EU (France)*

Magnetic confinement fusion is a promising pathway towards sustainable and basically inexhaustible energy source. Nuclear energy can be originated in fission of uranium nuclei (heaviest element in nature) or by fusion of hydrogen nuclei (lightest element in nature). All existing nuclear power plants are based on nuclear fission phenomenon, with the known drawbacks of uncontrolled ac-

cidents or long-life radioactive waste. Conversely, fusion energy will have no long-life radioactive waste and are inherently safe: the harmless Helium gas is the only product of the nuclear reactions.

For decades, Europe has been at the forefront of fusion research with the construction and operation of tokamaks and stellarators, with large-scale projects such as the Joint European Torus (JET), and other electromagnetic fusion devices such as Tore Supra and WEST in France, ASDEX and W7-X in Germany, TJ-II in Spain, DTT in Italy and JT-60SA in Japan most of them EU publicly funded devices. On top of this, EURATOM through Fusion for Energy (F4E) is sharing 45% of ITER, a fusion reactor, under construction in the south of France, whose objective is to demonstrate the scientific and technological feasibility of fusion energy.

In parallel, European national programmes and private initiatives are accelerating progress in advanced tokamaks, stellarators, and alternative confinement approaches. Challenges remain, including optimizing plasma stability, improving material resilience, and scaling up energy-efficient technologies for commercial deployment.

Looking forward, the European Commission is considering its fusion strategy with the objective of accelerating fusion commercialisation (installations connected to the grid). With sustained investment and international collaboration, magnetic confinement fusion could become a cornerstone of Europe's carbon-free energy strategy in the coming decades.

IV.3 The Present and Future of Inertial Fusion Energy (IFE) in Europe

Invited contribution

Markus Roth, `markus@focused-energy.co`*, Focused Energy GmbH & TU Darmstadt, Germany*

The scientific breakthroughs in fusion ignition and burn, driven by lasers, have accelerated research and development towards a first commercial fusion powerplant world-wide. While most of the scientific work has started in the Inertial Confinement Fusion (ICF) environment in the US the prospect for European Inertial Fusion Energy (IFE) progress is large.

Fostering the research mainly done in the UK and France, leveraging of European earlier initiatives like HIPER, a number of IFE Startups have evolved and are pursuing fusion energy concepts.

I briefly summarize the history and the current status of IFE research in Europe and show current initiatives around the recent fusion boost in Germany. I present a potential roadmap and facility layout required to achieve commercial fusion until the end of the 2030's.

IV.4 AI for Enhancing the Integration of Renewable Energies in Modern Power Systems

Invited contribution

Georges Kariniotakis, georges.kariniotakis@minesparis.psl.eu*, Mines Paris–PSL, France*

Modern power systems face challenges from the integration of renewables energy sources (RES) and other technologies like storage, demand response, electric vehicles, or power to X. The increasing complexity of these systems requires an intelligent layer for secure, economic, and resilient power system management. Artificial Intelligence (AI) already supports key functions like forecasting, scheduling, congestion management, predictive maintenance and others. Among these, load and renewable energy source (RES) forecasting tools are the most mature and widely used. Wind and solar production are highly variable and weather-dependent, creating uncertainties for system operators. Short-term forecasting (minutes to days ahead) is essential for secure operation and efficient market participation. Despite decades of progress, large forecast errors still occur, with significant technical and financial impacts. Forecasting has evolved from statistical and physical models to hybrid and AI-based approaches, including probabilistic methods. The European H2020 Smart4RES project advanced RES forecasting research through a holistic approach covering all the value and model chain of forecasting, from data to applications. In this presentation we highlight innovative research directions based on AI, that include seamless forecasting across all time horizons, resilient models to missing data, and value-oriented forecasting. Decision-focused learning and interpretable AI are explored. Privacy-preserving data sharing frameworks were proposed to leverage distributed data in forecasting and optimization functions while respecting confidentiality. Key takeaways stress the need for simplicity, trustworthy AI models, fair benchmarks, open data, and replicability. Overall, AI offers new opportunities, but must balance accuracy, robustness, and practical value in real-world applications.

IV.5 Data-Driven Methods for Integrated Energy and Transportation Networks

Invited contribution

Bissan Ghaddar, bghaddar@ivey.ca*, DTU, Denmark & Ivey Business School at Western University, Canada*

The integration of energy and transportation networks presents unique challenges and opportunities, driven by the growing demand for electric vehicles, renewable energy sources, and intelligent infrastructure. Optimizing these interconnected systems requires innovative methodologies that balance efficiency,

sustainability, and resilience. This talk highlights advanced optimization methods, including mathematical modeling and uncertainty quantification, to address the complex interactions between energy generation, distribution, and transportation systems. Key applications include optimizing electric vehicle charging networks for last-mile delivery, electrifying public transport, autonomous shuttle-sharing network design, and the deployment of shared energy storage systems. Machine learning and data-driven approaches are emphasized as tools to enhance predictive modeling, enable real-time decision-making, and manage the inherent complexity of these networks. We provide insights into integrating machine learning techniques, such as reinforcement learning and data-driven methods, with traditional optimization frameworks to achieve robust and scalable solutions. By leveraging digital technology, advanced computational tools, and societal considerations, the goal is to design intelligent, sustainable, and resilient energy and transportation systems that address the future challenges while meeting the growing demand of the population.

Chapter V

Digital Regulation and Strategy in the EU

V.1 Digital Strategy in the EU: Lessons from AI Innovation in China

Invited contribution
Claudio Feijóo, `claudio.feijoo@upm.es`*, Universidad Politécnica de Madrid (UPM), Spain*

The talk departs from the acknowledgement of a virtuous cycle that often emerges between disruptive innovations and industrial leadership, as pioneering breakthroughs accelerate competitiveness and open new commercial frontiers. In this context, and looking for valuable lessons of interest for Europe from recent advancements in China, artificial general intelligence (AGI) has become the focal point of global rivalries, with major economies vying to secure an edge in both foundational research and industrial deployment. To this regard, China has proven that incremental innovations play a key role in solidifying a country's industrial base and market dominance. At the same time, recent restrictions on advanced AI chip exports have underscored the dependencies and strategic vulnerabilities of AI supply chains, prompting countries to invest in more resilient ecosystems but not necessarily detaining innovation, rather the contrary, accelerating it. This shifting landscape also reveals how essential local hyperscalers can be, not merely to handle training complexities but to ensure data sovereignty and robust AI infrastructure.

The talk also briefly explains how the pressing challenges in optimizing training for large language models (LLMs) is moving towards inference as a crucial bottleneck in advancing AI capabilities, as more robust reasoning models, AI agents, and automated AI research demand efficient, scalable inference solutions and which are the lessons for Europe. Also, as new AI paths-like spatial

intelligence-gain traction, novel avenues appear for enhancing both real-world applications (e.g., geospatial analytics) and next-generation digital services. Ultimately, Europe must learn from China's experience in bridging research breakthroughs, industrial pragmatism, and capacity-building, particularly as all of us grapple with the most profound dilemma yet: how to contend with human-level AI potentially emerging by 2030.

V.2 Empathic AI Companions: Navigating Ethics, Risks, and Public Perceptions

Invited contribution
Andrew McStay, mcstay@bangor.ac.uk*, Bangor University, United Kingdom*

This talk explores public attitudes toward AI companions, examining the perceived benefits, concerns, and governance preferences surrounding these emerging technologies. Motivated by recent controversies, including AI's involvement in emotionally sensitive contexts like mental health and companionship, this discussion critically evaluates AI companions' potential to alleviate loneliness, enhance intimacy, and support vulnerable groups such as children and older adults. Simultaneously, it highlights significant ethical challenges, including emotional dependency, deception through anthropomorphism, and risks of exploitation.

Drawing on UK public survey data ($n = 2073$), this talk contextualizes these technologies within broader societal implications, including their impact on human relationships and psychological well-being. It examines governance strategies, addressing whether existing regulatory frameworks suffice or if new provisions are necessary to mitigate risks. Additionally, the talk considers innovative perspectives on AI as relational entities, integrating ideas from global ethical traditions that challenge anthropocentric notions of community and personhood.

Concluding with recommendations, the talk advocates for hard law, but also soft law guidance on inclusive AI design that prioritizes user well-being, safeguards against harm, and integration of public voices to shape a safer generation of AI companions.

V.3 Digital Identity, the Specific Case of Age Verification

Invited contribution
Olivier Blazy, olivier.blazy@polytechnique.edu*, École Polytechnique – IP Paris, France*

We explore innovative approaches to age verification that prioritize user privacy and data sovereignty. Traditional age verification methods often require individuals to disclose personal information, leading to potential privacy concerns and data misuse. Modern policies advocates for systems that enable users protect their privacy without revealing additional personal data, while ensuring their statement is valid thereby enhancing both security and user autonomy.

A notable example of such an approach is the privacy-preserving age verification system developed in collaboration with the LINC[1] and the *Pôle d'Expertise de la Régulation Numérique* (PEReN). This system allows users to access age-restricted websites without sharing personally identifiable information, ensuring compliance while maintaining user privacy.

We also discuss the interplay between these privacy-centric verification methods and the forthcoming eIDAS 2.0 regulation. The eIDAS 2.0 framework aims to provide EU citizens with digital identity wallets, enabling them to prove their identity and share electronic documents securely across the EU. These wallets are designed to allow users to disclose only specific personal attributes, such as age, without revealing their full identity, thereby granting individuals greater control over their personal data.

By integrating privacy-preserving age verification methods with the capabilities of eIDAS 2.0 digital identity wallets, citizens can regain ownership of their data and privacy while accessing various services. This synergy not only enhances user freedom but also aligns with the broader goals of digital sovereignty and trust in the digital ecosystem.

Through the lens of the French age verification legislation, and the eIDAS 2.0 framework, we show the importance of developing and implementing identity management systems that respect user privacy. By leveraging frameworks like eIDAS 2.0, it is possible to create a digital environment where citizens can access services securely and privately, maintaining control over their personal information.

V.4 "Free" AI Models as Software: Freedoms and Rights in the EU AI Act

Regular contribution

Simona Ramos (speaker, `simonaramos95@gmail.com`*), UPF Barcelona School of Management (BSM), Spain; joint work with Fabio Pianese, Nokia Bell Labs, France*

The EU AI Act attempts to mitigate systemic risk with the provisioning of AI services on the European Union's internal market [1]. In an effort to promote innovation and technological development, Article 53(2) establishes exceptions on certain obligations, including reporting requirements, that deployers of AI services that are not deemed as high-risk need to fulfill based on the models

[1]Digital Innovation Lab (*Laboratoire d'Innovation Numérique*) of the French National Commission on Informatics and Liberty (CNIL).

being released under a "free and open-source licence".

The policy debate on the purpose of and eligibility to such an exception is still very much active, as without its clear definition, the EU risks encouraging superficial compliance that undermines the Act's broader objectives and societal impact. To fully capture the implications of freedom in AI development, a deeper philosophical and legal exploration is needed, where freedom is understood not just as the public availability of partial components of an AI system, but as the empowerment of the public to study, modify, and share openly AI models, exercising their agency within a system that values transparency and fairness. Recent related work [2] sets out to precisely characterize the "openness" of AI models that may qualify for the exception under the AI Act, to prevent the threat of "open-washing" practices that could trigger the exception by only disclosing minimal information on superficially open models. Although carefully investigating the "open-source" angle, the authors' analysis does not dwell on the "free" adjective in the text of the AI Act. But the distinction is important, as models such as META's Llama 3.1, whose license is touted as open and permissive, do not qualify as free under more careful scrutiny [3]. Our contribution aims to justify a stricter understanding of the "free and open-source" exception.

The explicit AI Act reference to "free licenses" urges us to explore the radical analogy that exists between the writing of a program and the training of a model, coupled with the public interest in which the exception of Article 52(3) needs to be grounded. The visible objective of the exception is in fact to foster innovation by exempting from most reporting requirements, save for the literacy (Art. 4) and transparency (Art. 50) obligations, the model deployers (who are not providers themselves) who employ "free and open-source" models for low-risk AI-based services [4]. At the same time, the AI Act chiefly regulates entities that place AI systems on the market and/or deploy them as services ("providers", "deployers", "importers", "distributors") (Articles 6-25). Models that are simply made available for download under a non-commercial license, without a party responsible for their sale or deployment as a service, appear to fall outside of the direct regulation. On the other hand, according to the definition by the Free Software Foundation, a program is free if its users are granted four essential freedoms [5]. While an end goal in itself for open-source software, the right of access to code is simply a necessary but non-sufficient prerequisite for the user freedoms to be exercised, and to this end some obligations are placed on both service providers and software distributors, regardless of the commercial character of the software or service.

Let us suppose that the original developer of an AI model, acting neither as a provider nor as a distributor according to the AI Act, makes a model available for download to the wider public while concealing as legitimate trade secrets those essential parts that would make it "free", such as the training code or the training dataset and its full attribution to the original contributors. An exception granted to deployers on grounds of public utility would therefore become detrimental to the right of the users to understand and verify the behavior of the service, including their ability to check for artifacts that might arise from the use during training of public domain datasets (or datasets for which consent was

acquired). Ultimately, the lack of access to training data is a major obstacle to the verification of the compliance to the low-risk classification of a service. The Article 53(2) exception should be made to apply to AI models that developers release under free software terms, ensuring that both deployers and users can audit models, understand harms, and uphold ethical standards. Without this, the exception becomes a shield for opacity, enabling unchecked proliferation of gray-box systems that claim openness while obscuring their data lineage [6]. Therefore, defining "free and open-source" becomes a question of accountability [7]. Only by demanding accountability at the source –not just at deployment– the AI Act can preserve the spirit of open-source innovation while safeguarding civil liberties and agency over AI outcomes.

A key unresolved issue underlying the EU AI Act is the arbitrary perimeter of AI concept in the regulation, which does not attempt to provide a technical definition of "what AI is or could be", as [1] observes. Thinking of AI as software and referring to the vast literature on software issues might help clarify certain conundrums, such as the "free and open" exception, and hopefully provide further guidance in the application of the upcoming regulation.

References

1 Ruschemeier H. AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. ERA Forum. 2022;23:495-510.

2 Liesenfeld A, Dingemanse M. Rethinking Open Source Generative AI: Open-Washing and the EU AI Act. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT'24). ACM; 2024. p. 1774-84. Available from: https://dl.acm.org/doi/10.1145/3630106.3659005.

3 Siewicz K. Llama 3.1 Community License is not a free software license; 2025. Available from: www.fsf.org/blogs/licensing/llama-3-1-community-license-is-not-a-free-software-license.

4 Fazlioglu M. EU AI Act Compliance Matrix; 2025. Available from: https://iapp.org/media/pdf/resource_center/eu_ai_act_compliance_matrix.pdf.

5 Stallman RM. What is Free Software?. Free Software Foundation; 1996–. Available from: https://www.gnu.org/philosophy/free-sw.html.

6 Lawson-Hetchely C. The Potential Impact of the Future AI Act on the GDPR; 2022. Available from: https://duo.uio.no/handle/10852/100000.

7 Novelli C, Taddeo M, Floridi L. Accountability in artificial intelligence: what it is and how it works. AI Society. 2024;39:1871-82.

V.5 Survive or Thrive? The AI Act and its Legal Impact in Business

Invited contribution

Neringa Gaubienė, neringa.gaubiene@tf.vu.lt*, Vilnius University, Lithuania*

The AI Act represents a world-first regulatory effort by the European Union to govern artificial intelligence, aiming to ensure transparency, fairness, and accountability. However, as Emmanuel Macron aptly warned, Europe is "over-regulating and under-investing" [1]. If the current trajectory remains unchanged, the EU risks becoming a "cheap tech talent pool" for the US, rather than a global leader in AI innovation. This paper examines whether the AI Act is a catalyst for responsible AI adoption or a barrier that could push European businesses out of the market.

The core challenge of AI regulation is balancing innovation and control – after all, as the saying goes, "cars drive faster with brakes". Regulation, like brakes, should enhance security without stifling progress. Yet, businesses face increasing compliance burdens, particularly for high-risk AI systems. As David Rogers, a leading author on digital transformation [2], argues, "Digital transformation should be business, not tech-driven". The AI Act, however, places legal obligations before business realities, demanding extensive documentation, compliance assessments, and risk management processes that disproportionately burden startups and SMEs.

A fundamental issue often overlooked in AI regulation is data quality. The principle of "junk in, junk out" remains critical: AI is only as good as the data it is trained on. While computing power continues to grow exponentially, data –the fossil fuel of AI– is reaching its limit. Businesses that want to harness AI's full potential must invest in high-quality, structured datasets. Those that lack such data must start collecting and refining it now, or risk falling behind.

The Brussels Effect –where EU regulations influence global markets beyond its borders– is expected to extend to AI, shaping global compliance standards. This extraterritorial impact means that companies worldwide may be forced to align with EU requirements, even if they operate outside Europe. While this could position the EU as a global regulatory leader, it may also push AI companies to develop products elsewhere, in jurisdictions with fewer restrictions and stronger investment incentives. If businesses perceive EU regulations as overly restrictive, they may relocate R&D to other jurisdictions, where AI innovation faces fewer legal roadblocks and greater financial support.

The AI Act could set a global benchmark for AI governance, but without significant investment, strategic flexibility, and business-first thinking, Europe risks regulating itself out of the AI race. If regulation continues to outpace innovation, European businesses will struggle to thrive, rather than simply survive, in the AI era.

References

1 Emmanuel Macron, "Europe Speech". April 25th, 2024.
Available at: `https://www.elysee.fr/en/emmanuel-macron/2024/04/24/europe-speech`

2 David L. Rogers, "The Digital Transformation Playbook: Rethink your Business in the Digital Age". Columbia Business School Publishing. NYC: 2016.

V.6 Noise Injection Reveals Hidden Capabilities of Sandbagging Language Models

Abstract accepted for inclusion in proceedings
Philipp Kreer (philipp.a.kreer@tum.de), joint work with Cameron Tice, Nathan Helm-Burger, Prithviraj Shahani, Fedor Ryzhenkov, Jacob Haimes, Felix Hofstaetter, Teun Van der Weij, TU Munich, Germany

Capability evaluations are critical in ensuring the safe deployment of frontier AI systems and enforcing regulations like the AI Act. Intentionally underperforming AIs, known as sandbagging, undermine the pivotal role of capability evaluations. Frontier multimodal Language Models like GPT o1 or Claude Sonnet-3.5 can recognize evaluation environments and underperform to avoid shutdown or regulations. Moreover, password-locked models conceal their capabilities unless the user inserts the correct password into the prompt.

I present a novel model-agnostic method for detecting sandbagging AIs. Injecting Gaussian noise into the parameters of sandbagging models disrupts the deceptive behavior, yielding an anomalous and considerable performance improvement. I demonstrate the effectiveness of noise injection for various open-weight large language models on the Weapons of Mass Destruction Proxy benchmark, which evaluates hazardous knowledge in biosecurity, cybersecurity, and chemical security.

Frontier labs or regulatory bodies can immediately apply noise injection to improve the trustworthiness of capability evaluations.

Chapter VI

Data Platforms, AI Technologies, and Regulation

VI.1 Digital Rights, Personal Data Collection and Inferences: the Case of Gender

Invited contribution
Gloria González-Fuster, Gloria.Gonzalez.Fuster@vub.be*, VUB, Belgium*

Data about gender is regularly processed for a variety of purposes, and in many different ways. Sometimes it is directly collected from data subjects, whereas in some other cases gender is inferred, again potentially in a variety of ways, ranging from simple inferences based on a title provided by the individuals, or based on profiling of the individual's behaviour. This diversity of processing methods has an impact on the manner in which the law regulates the processing of personal data, and on the rights that data subjects may use to access, rectify, contest, or request the erasure of such data. Contrary to popular belief, gender data is not regarded by the General Data Protection Regulation (GDPR) as 'sensitive data', although its potential for discrimination has been acknowledged and, depending on the context, it might need to be considered as revealing (other) 'sensitive data'. This contribution looks into existing practices and the most recent case law of the Court of Justice of the European Union (CJEU) to explore the intersection between digital rights and gender data.

VI.2 Regulating Dark Patterns in the EU: Legal Challenges and Policy Responses

Regular contribution
Pratiksha Ashok, P.Ashok@tilburguniversity.edu*, Tilburg University, The Nether-*

lands

The rapid digitalization of commerce and communication has transformed how consumers interact with goods and services. However, alongside the benefits of digital innovation, new risks have emerged, particularly in the form of *dark patterns*. These are deceptive design strategies embedded within digital interfaces that manipulate users into making choices they might not otherwise make. These patterns exploit cognitive biases and behavioural tendencies to steer users toward actions that primarily benefit businesses, often at the expense of consumer autonomy and welfare. Examples of dark patterns include false urgency cues, forced consent to data sharing, hidden fees revealed late in the purchasing process, and complicated processes to cancel subscriptions. While the European Union (EU) has recognized the dangers posed by such manipulative tactics and has developed a framework of laws to protect consumers, the absence of a clear, unified definition of dark patterns and the fragmented nature of enforcement mechanisms continue to undermine effective regulation and consumer protection. This essay explores the concept of dark patterns, the current EU legal framework addressing them, the enforcement challenges faced, and the urgent need for clearer, more cohesive regulatory responses to protect consumers in the digital age.

Firstly, it is essential to understand what dark patterns are and why they are problematic. The term "dark patterns", coined by user experience designer Harry Brignull, refers to interface designs that intentionally mislead or coerce users into decisions that they might not make if fully informed or given a fair choice. These manipulative designs capitalize on psychological principles such as scarcity, social proof, or default bias to influence user behaviour. For instance, a website might display a countdown timer suggesting limited availability to create a false sense of urgency, prompting hurried purchases. Similarly, some platforms require users to consent to unnecessary data processing as a condition for accessing a service, effectively forcing action through design. Other dark patterns include hiding additional costs until the final step of a transaction or making the process of unsubscribing from a service deliberately complex and time-consuming. These practices are not merely annoying and distracting; they can cause significant harm to consumers, including financial loss, unwanted data exposure, and erosion of trust in digital services. Vulnerable groups, such as minors or individuals with cognitive impairments, are particularly susceptible to these manipulative tactics, raising important ethical and legal concerns.

The EU has responded to deceptive digital practices through a variety of legislative instruments, though none explicitly define or comprehensively regulate dark patterns as a distinct category. Instead, the regulatory approach relies on interpreting broader provisions related to unfair commercial practices, transparency, and data protection. The Unfair Commercial Practices Directive (UCPD) prohibits misleading actions and omissions in commercial communications, which can encompass certain dark patterns such as false urgency or hidden costs.

The Consumer Rights Directive (CRD) complements this by mandating clear

and transparent information in distance contracts, aiming to ensure consumers are adequately informed before concluding online transactions. The General Data Protection Regulation (GDPR), effective since 2018, addresses manipulative consent mechanisms by requiring that consent for data processing be freely given, specific, informed, and unambiguous. This has particular relevance for dark patterns that seek to obtain personal data through deceptive or coercive means.

More recently, the Omnibus Directive strengthened transparency requirements in online marketplaces and targeted practices such as pre-ticked boxes and misleading subscription traps. The Digital Markets Act (DMA), adopted in 2022, imposes obligations on gatekeeper platforms to ensure, among other things, that users can easily withdraw consent and are not subjected to manipulative interface designs. The Digital Services Act (DSA) explicitly prohibits the use of deceptive interface designs that manipulate users, marking a significant step toward addressing dark patterns more directly. Additionally, the forthcoming AI Act introduces prohibitions on manipulative or exploitative AI techniques, particularly those targeting vulnerable individuals, further expanding the regulatory scope over emerging digital manipulation methods.

The proposed Digital Fairness Act (DFA) represents a significant step forward in the EU's regulatory response to dark patterns by introducing a broader and more cohesive framework specifically designed to address manipulative digital practices. Unlike earlier legislation, which often addressed dark patterns only indirectly or within limited contexts, the DFA applies to a wide range of digital businesses including e-commerce platforms, app developers, digital advertisers, and online service providers, and adopts a more expansive definition of manipulative design. The Act explicitly targets deceptive interface techniques that exploit cognitive biases and psychological vulnerabilities, such as making subscription cancellations unnecessarily difficult or leveraging AI-driven profiling to nudge users toward unintended actions. Additionally, the DFA introduces stricter transparency requirements for influencer marketing and algorithm-driven promotions, aiming to curb misleading commercial practices that have proliferated in the digital marketplace. By focusing on "fairness by design", the DFA seeks to ensure that digital interfaces are structured to support informed and autonomous decision-making, rather than to manipulate or deceive users. This comprehensive approach reflects a growing recognition among EU policymakers that fragmented and piecemeal regulation is insufficient to protect consumers from the evolving risks posed by dark patterns, and that a unified legal standard is essential for achieving genuine digital fairness in the online environment.

Despite this extensive legislative framework, a critical gap remains, i.e., there is no unified or explicit legal definition of dark patterns within EU law. This absence creates significant challenges for both businesses and regulators. Without a clear definition, companies face uncertainty about which design practices are lawful, potentially chilling innovation or allowing harmful tactics to persist unchecked. For regulators, the lack of clarity complicates enforcement efforts, as authorities must interpret general provisions on unfairness or deception on

a case-by-case basis, leading to inconsistent outcomes across Member States. This legal ambiguity undermines the goal of harmonized consumer protection within the EU's single market and leaves consumers vulnerable to manipulative digital experiences.

Enforcement of rules related to dark patterns is further complicated by the fragmented implementation in the EU. Multiple authorities share responsibility depending on the legal instrument and sector involved. Consumer protection agencies enforce provisions under the UCPD and the CRD, data protection authorities oversee GDPR compliance, competition authorities monitor gatekeeper behaviour under the DMA, and sectoral regulators handle specific domains such as financial services or advertising. This multiplicity of regulators can lead to overlapping or conflicting decisions, regulatory uncertainty, and uneven enforcement. Moreover, businesses operating across different Member States face divergent interpretations and enforcement practices, increasing compliance costs and complicating cross-border digital commerce.

Another enforcement challenge stems from the interpretive complexity of dark patterns. Many manipulative design choices exist on a spectrum between legitimate marketing and unlawful deception. For example, countdown timers may sometimes reflect genuine scarcity, while in other cases, they create false urgency. Similarly, the placement and colour of buttons or the wording of consent requests may be persuasive but not necessarily manipulative. Regulators must carefully assess the context, intent, and actual impact of these design elements to distinguish between acceptable persuasion and unlawful manipulation. This requires a nuanced, principles-based approach rather than rigid, prescriptive rules, to avoid stifling legitimate business practices while effectively targeting harmful ones.

Compounding these difficulties is the *rational actor's assumption* embedded in much of EU consumer law, which presumes that consumers make decisions based on rational evaluation of information. Dark patterns exploit cognitive biases and emotional triggers, undermining this assumption and rendering traditional legal standards less effective. For instance, consumers may act impulsively under the influence of false urgency or social pressure, leading to unintended commitments or data disclosures. This behavioural insight gap calls for updated legal standards that recognize and address the realities of consumer decision-making in digital environments.

Academic commentary and stakeholder input have highlighted these challenges. Legal scholars emphasize the inconsistencies and under-enforcement resulting from fragmented frameworks and ambiguous definitions. They advocate for legal updates that can keep pace with emerging dark patterns, including AI-driven hyper-nudging and personalized manipulation. Consumer organizations and industry stakeholders alike have called for clearer, stronger safeguards against dark patterns, comprehensive bans on the most harmful tactics, and harmonized enforcement to reduce uncertainty and foster trust.

The debate on how best to regulate dark patterns includes divergent views. Some argue that the existing EU legal framework is sufficiently robust if properly enforced and that introducing additional laws may increase complexity without

delivering meaningful improvements. They recommend clarifying the interaction between existing rules, issuing harmonized guidelines for enforcement and compliance, enhancing cooperation among national authorities, and focusing on intent and harm rather than prescriptive design bans. This approach aims to balance consumer protection with business innovation and flexibility.

Others contend that the persistent gaps and enforcement challenges necessitate more substantive regulatory action. They call for the adoption of a *common, technology-neutral definition of dark patterns* to provide legal clarity and certainty. Binding or authoritative guidelines could assist regulators and businesses in identifying and evaluating manipulative designs consistently. Establishing coordinated enforcement mechanisms, possibly through enhanced cooperation or a centralized EU body, would help ensure uniform application of rules across Member States. Additionally, regulatory frameworks must be adaptable to technological advances, including AI and data-driven personalization, to remain effective against evolving manipulation techniques.

Special attention is warranted for vulnerable groups, such as children and cognitively impaired individuals, who are disproportionately affected by dark patterns. The law should incorporate heightened protections for these users, recognizing their limited capacity to resist manipulative designs and the greater harm they may suffer.

Ultimately, the EU's regulatory approach to dark patterns should strive for digital fairness, ensuring that digital markets operate transparently and ethically, respecting consumer autonomy and fostering trust. This requires a cohesive, principles-based legal framework that balances innovation with protection, is adaptable to technological change, and is supported by clear enforcement mechanisms.

In conclusion, dark patterns represent a significant and complex challenge for consumer protection in the EU's digital economy. While the existing legislative framework comprising the UCPD, Consumer Rights Directive, GDPR, Omnibus Directive, DMA, DSA, and forthcoming AI Act provides important tools, the lack of a unified definition and fragmented enforcement undermines their effectiveness. Addressing these shortcomings requires a concerted effort to clarify legal standards, harmonize enforcement, and develop future-proof regulations that recognize behavioral realities and protect vulnerable users. By doing so, the EU can ensure that digital innovation benefits consumers without enabling exploitation, thereby reinforcing trust and fairness in the digital single market.

References

1 Gray CM and others, 'The Dark (Patterns) Side of UX Design', Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (ACM 2018) `https://dl.acm.org/doi/10.1145/3173574.3174108`, accessed 11 January 2025

2 Mathur A and others, 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites' (2019) 3 Proceedings of the ACM on Human-Computer

Interaction 1

3 Polona Car and Filippo Cassetti, Members' Research Service, 'Regulating Dark Patterns in the EU: Towards Digital Fairness' (EPRS - European Parliamentary Research Service)

4 'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA Relevance.'

5 Rieger S and Sinders C, 'Dark Patterns: Regulating Digital Design'

6 Santos C, Morozovaite V and De Conca S, 'No Harm No Foul: How Harms Caused by Dark Patterns Are Conceptualised and Tackled under EU Data Protection, Consumer and Competition Laws' (2024) `https://www.ssrn.com/abstract=4877439`, accessed 13 January 2025

7 The Legal Consistency of Technology Regulation in Europe (Hart Publishing 2024)

8 Trzaskowski J, 'Persuasion, Manipulation, Choice Architecture and Dark Patterns' in Andrej Savin and Jan Trzaskowski (eds), Research Handbook on EU Internet Law (Edward Elgar Publishing 2023) `https://www.elgaronline.com/view/book/9781803920887/book-part-9781803920887-25.xml`, accessed 13 January 2025

9 Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts (UCTD) (OJ L 95, 2141993)

10 Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive') (OJ L 149, 1162005)

11 Regulation 2016/679/EU on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016 (OJ L 119, 452016)

12 Regulation 2022/868 on European Data Governance and amending regulation 2018/1724 (Data Governance Act) 2016 (PE/85/2021/REV/1 OJ L 152, 362022)

13 Regulation 2022/1925 on contestable and fair markets in the digital sector and amending Directive EU 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (PE/17/2022/REV/1 OJ L 265, 12102022)

14 Regulation (EU) 2022/2065 on a Single Market for Digital Services and amending directive 2000/31/EC (Digital Services Act 2016 (PE/30/2022/REV/1 OJ L 277, 27102022)

VI.3 Legal Challenges of Text and Data Mining in the Era of Generative AI

Invited contribution

Artūras Grumulaitis, arturas.grumulaitis@tf.stud.vu.lt*, Vilnius University, Lithuania*

Ability of Generative AI (GenAI) to produce realistic and high-quality outputs makes it a transformative tool in many industries, including technology, entertainment, healthcare, and education. This technology relies on vast amounts of data, which must be carefully collected, prepared, and processed to function effectively. Developers of GenAI models should ensure that training data is sourced legally, with proper authorization from intellectual property (IP) holders and fair compensation where required. However, in practice, training datasets are often built from publicly available Internet sources, including copyrighted literary, scientific, and artistic works. The inclusion of such works without explicit permission raises important legal concerns, particularly regarding copyright protection.

The EU AI Act (Article 51.1(c)) establishes that providers of general-purpose AI models must "put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790". Accordingly, the Directive (EU) 2019/790 on Copyright in the Digital Single Market (DSM Directive) introduces two exceptions for text and data mining (TDM), allowing researchers and businesses to use copyrighted materials under certain conditions. However, key aspects of these exceptions remain unclear when applied to GenAI. Questions persist about the scope of permitted use, access to the training data, and how rights holders can opt out of dataset inclusion. Recent legal cases, such as Kneschke v. LAION, highlight the ongoing debate over whether using copyrighted images for AI training falls within the scope of lawful TDM.

This research examines the legal challenges surrounding TDM and copyright, focusing on the balance between AI innovation and intellectual property rights. It explores how legal frameworks are evolving to address these issues and what this means for AI governance. As AI systems continue to rely on large datasets, achieving legal clarity is essential to support technological progress.

The study concludes with recommendations for policymakers, researchers, and industry stakeholders on navigating the complex legal landscape of TDM. By addressing these challenges, the AI community can work towards a responsible and legally sound approach to AI development.

VI.4 Trust, Disinformation, Verification: Issues, Approaches, Challenges

Invited contribution Jochen Spangenberg, jochen.spangenberg@dw.com*, Deutsche Welle (DW), Germany*

(This is an edited version of the speech by Jochen Spangenberg, at the EUDTP conference held in the European Parliament, Brussels, on 15 May 2025)

To start with, allow me to provide you with some context of a more general nature. I feel this is important for the sake of transparency and for the reader to understand a bit better where I am coming from.

I consider this of particular importance as we are dealing with topics such as

- disinformation,
- polarization,
- independence,
- impartiality,
- bias,
- truth,
- trust,
- democracy,
- the role of media (old and new),

and such like.

I have been working in the media sector all my professional life (around 30 years by now) and studied media and communication sciences, which I still teach as a Visiting Lecturer, primarily at the Free University Berlin, whenever time allows.

Of those 30 odd years in the media sector, around 25 were spent working in and for public service media organisations. This is partly due to a bit of luck and circumstances, but also has to do with a conviction and a belief that I have held since my twenties: namely that there are very important reasons public service media exist, what they stand for, and what they try to achieve. That is: if you are a fan of democracy and free and open, pluralistic societies, of course.

Don't get me wrong: I am not for a minute arguing that everything we have with regards to public service media or individual organisations is perfect. Far from it!

However, I still think that, even in today's digital world, media organisations that are not driven primarily by commercial imperatives have an important role to play. By this I mean organisations that are driven by other aspects or forces than purely by maximizing revenue, such as

- work ethics,
- a certain view of society and their audiences (namely regarding them as citizens rather than consumers), and
- a particular approach to what is covered and how this is done.

Now as I work for one such organization –and with all I said beforehand– you know how to take what is coming next to some degree. I also feel this is important for the sake of transparency, something also of relevance in this topical context.

Next, I will zoom in on the individual constituent parts of my topic which are: "Trust, disinformation and verification". I will address them one by one. Let me start with *trust*.

Trust is something that is very hard to earn, but easy to lose!

If you work in news media, you have to earn it again and again, day-in, day-out. I still remember vividly when, in my 20s, a BBC colleague once said something along the lines of the following: "We get it right a thousand times, and nobody gives a toss and quickly forgets about it. That's what people expect. But when we get it wrong once, everyone remembers. That's why our aim should always be to try our very best to get it right!"

Not an easy task. But worthwhile, nevertheless. A challenge for sure in today's fast-paced news cycles. All in a world in which, at least in theory and technically, everyone can be an "information disseminator".

Trust is one of our primary assets – and here I mean in particular news media organisations that aim to report fairly, transparently, independently and impartially. I would even say this is the foundation of our "business model" – and I am referring to public service media organisations in particular.

What we must not forget is that there are various forces and actors that try and do exactly the opposite, or better: they want to undermine media that strive to be as independent and impartial as possible. Because these actors have their own agendas. These agendas can be political in nature, or financially driven. We –as a society and individuals– need to be aware that this happens, and why there are some who try and undermine trust, or sow distrust. This is also why we need the support of various sectors and actors. Very high on the list are politicians and political parties – at least those that subscribe to democracy, freedom of the press and such like and believe in free and pluralistic societies. Because –and this is something we are seeing more and more– numerous actors have a profound interest in spreading something that I will cover next: namely the second item in the title of my talk: *disinformation*.

The emergence of social networks and developments in digital technology have resulted in fundamental changes for both the way information is spread as well as consumed. Traditional or legacy news media no longer function as gatekeepers. In other words: in the past, only a selected few had the means to disseminate information to large numbers of people. These days, everyone can be an information provider technically. All that is needed is a digital device and an online connection.

Further, if you have enough financial resources, you can buy or develop (or also destroy) certain platforms that allow for information dissemination in a new way. One prime example is Elon Musk and what he made of Twitter, turning it into toxic X – or his private mouthpiece.

When we look at so-called "social" networks generally (almost all of them commercially funded) it is not the truth or value of information that matter

most there. Other things count, such as content that polarizes, emotionalizes, and results in interaction in whichever way. In other words: anything that creates clicks and can be monetized.

Now it is important that not only experts are aware of these dynamics, but, ideally, citizens are, too. This requires a lot of work, especially in the form of education. Or to be more precise: media literacy.

With this as a snapshot allow me to move to the third and final component of my topic's headline: *verification*.

This is where not only legacy media come in, but also so-called fact-checking organisations, which have been established primarily over the past six to eight years or so as an almost new ecosystem. So let me move briefly to fact-checking.

One could argue that this is what journalists do all the time anyway. I agree to some extent. However, with advances in technology and the ever-increasing amount of disinformation, it is also important to "set things right" and correct what others are spreading if it is untrue. Especially when disinformation comes from people or channels with enormous reach and vested interests.

The downside: while fact-checking is vitally important these days, those who do it are always running behind in a way. They are not creating any new narratives or setting new topics, ideally in a constructive, solutions-oriented way. Instead, it's like a game of "whack-a-mole": you can never win. As soon as you debunked one story, the next one pops up. A tiring and at times rather frustrating task – but also an important one.

Another dimension I want to address while dealing with the topic of verification is technology itself. On the one hand, technological developments have played into the hands of those who are spreading disinformation. But technological advancements can also be used by those involved in fact-checking and verification. This implies learning to use respective tools that are out there, but also developing new tools and detection methods that debunk disinformation and false narratives in a reliable way. This is where the EU plays an important role. One that needs to be applauded.

I myself and the organisation I work for, Deutsche Welle, have been involved in a number of research & development projects that deal with disinformation detection and developing supporting tools and services. One such project, running from September 2022 until late in 2025 is called veraAI[1]. In my view, a very successful project.

My request and advice here is to keep such projects and initiatives and respective outcomes ongoing or have follow-ups, and make available adequate resources for such undertakings. This should be the case in the technology development domain, but also include making the topic a priority elsewhere. Additionally, the worlds of technology development and social sciences should be brought even closer together.

As I mentioned the EU, it also needs to be pointed out that numerous initiatives and activities of relevance have been initiated in the fight against disinformation by the EU already. This ranges from implementing the Digital

[1] Web: `https://www.veraai.eu/`

Services Act (DSA) and the Strengthened Code of Practice on Disinformation to supporting initiatives such as EDMO (the European Digital Media Observatory and the respective EDMO Hubs) and the EFCSN (European Factchecking Standards Network), to the work of the EEAS (European External Action Services) with regards to FIMI (Foreign Information Manipulation and Interference), and much more.

In principle –and looking at the world in its current state and related dynamics– I would argue that countering disinformation is more important than ever. Here, too, we are witnessing some kind of "cat and mouse game". Plus: with advances in Artificial Intelligence (from face swapping to audio manipulations at a user's fingertip), it is important that journalists and factcheckers can keep up as much as possible – and not get left too far behind.

To conclude briefly: there are numerous challenges and threats out there – that is if we can agree on the notion that guarding our democracies is something that unites us, as is the desire to live in free and pluralistic democratic societies. To protect and safeguard all this, or even let it prosper, requires efforts on a variety of levels. One such level or player are fact-checkers and news media organisations that strive to report as accurately as possible, being a constituent part of democracy. With respective technologies at hand, that is. In addition to checking and correcting disinformation, it also requires both resources and the backing to set own narratives and tell stories. Narratives that earn the trust of audiences and help them to stay informed and master living in an increasingly complex world. Audiences that are ultimately the ones who will decide about the future of media organisations and media systems, and with that our democracies. It is time to act and –I would argue– pick up speed and intensity, on numerous fronts. The reason is simple: a lot is at stake.

VI.5 Building the Next Generation of Social Media Research Tools: Implementation Challenges and Opportunities under the DSA

Invited contribution
Fabio Giglietto, `fabio.giglietto@uniurb.it`*, Università degli Studi di Urbino Carlo Bò, Italy*

This paper analyzes the evolving landscape of social media research APIs and tools under the European Union's Digital Services Act (DSA), with particular focus on Article 40.12's mandate for researcher access to publicly available platform data. Drawing on extensive testing of research APIs from major platforms including Meta, YouTube, and TikTok, we examine how Very Large Online Platforms (VLOPs) are implementing data access requirements. Our analysis reveals the critical need for more inclusive and comprehensive research tools that serve diverse stakeholders – from data scientists to qualitative researchers without coding expertise. We identify key implementation challenges

including inconsistent definitions of "publicly accessible data", restrictive rate limits, and technical barriers that limit research capabilities. The paper argues that the DSA represents a historic opportunity to develop standardized, well-documented APIs and user-friendly interfaces that could foster an ecosystem of reliable third-party research tools. We offer concrete recommendations for improving researcher access while ensuring tools remain accessible to users with varying technical expertise. Our findings contribute to ongoing discussions about platform accountability, research infrastructure development, and the future of social media analytics in the EU's regulatory framework.

VI.6 The 2024 EU AI Act's Approach to Stubborn AI: A Legal Perspective on Errors (Bugs) in Public Administration

Abstract accepted for inclusion in proceedings
Petia Guintchev, petia.guintchev@ub.edu*, Universidad de Barcelona (UB), Spain*

As Artificial Intelligence (AI) techniques develop, their integration within decision-making processes in both the private and the public sectors increases. Special relevance is ought to semi- and fully automated AI-powered processes and procedures within the public administration. Firstly, AI-powered public administration is not exempted of affecting the rights of the citizenry, particularly the right to good administration, enshrined in Art. 41 of the European Union's Charter of Fundamental Rights and recognized by domestic law of the EU Member States. Secondly, all software, whether involving Large Language Models (LLMs) specifically prone to hallucinations, Machine Learning (ML) systems which also showcase biases, or simpler digital solutions, present bugs. Some of these bugs hold significant legal relevance for administrative law. The negative impacts of bugs can hinder the lawful outcome of administrative procedures and impair the correct provision of public services, thus, hampering the possible benefits deriving from semi- or full automation of public administration, such as the personalization of public services or the efficiency and timeliness of administrative procedures.

This presentation focuses on the legal relevance of bugs within an AI-powered public administration by briefly referencing well-known problematic cases, including the French Affelnet case or the British Post Office case, and particularly the ongoing Spanish BOSCO case, which illustrate the significant difficulties in detecting and proving bugs due to several barriers: the lack of access to software documentation due to copyright protection, issues regarding explainability of the software and motivation of the semi- or fully automated administrative act, and the real effectiveness of responsibility and liability elucidating mechanisms. The study posits that the pernicious outcomes of errors and the difficulties they pose are primarily due to a lack of legal awareness regarding the need for a legal definition of bugs.

The lack of an explicit legal definition of bugs is evident in the regulatory efforts, particularly within the European context. For instance, the Counsel of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, and the Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (the EU AI Act), both adopted in 2024, pay scant, if any, attention to software/AI errors. Preparatory documents, academic literature and jurisprudential sources regarding bugs in AI-powered public administration are also scarce. Attention to a legal definition of bugs and corresponding redressing mechanisms is particularly important considering the aim of the EU towards a 100% digitalization of key public services, as well as the EU's Brussels effect. In this scenario, other concepts such as trustworthiness and robustness and accuracy, pivotal in the EU AI Act, are analyzed to assess whether they enclose an implicit, yet useful legal definition of AI errors.

Additionally, the wording of the EU AI Act and certain sectors of the legal scholarship prompted a discussion on the potential distinction between *simple* software and AI. This discussion is concatenating a series of arguments leading to suggest that the dichotomy, although useful in technology settings, may be artificial if enhancing the protection of citizens' rights and bolstering trust in AI-powered public administration is considered from a legal perspective. Other complementary strategies to legally address bugs from administrative law (analog categories of errors; the rule as code; the reserve of humankind; algorithmic transparency; taxonomies; human supervision; tolerable margin of error; or the right to the error) are also considered. However, these approaches present their own pitfalls.

All of the above reinforces the conclusion that a legal definition of software/AI errors will best serve the objective of achieving a trustworthy AI by fostering legal certainty and providing the pertinent redress mechanisms, which will ensure that AI's stubbornness, or persistent errors, are effectively addressed.

VI.7 The Impact of EU Digital Regulation on the Protection of Vulnerability Disclosure Researchers

Contribution accepted for inclusion in proceedings
Teodora Curelariu, `tcurelariu@gmail.com`*, Université Grenoble Alpes & INRIA, France*

Introduction

As the European Union accelerates its digital transformation, it has undertaken an ambitious legislative agenda aimed at strengthening cybersecurity. Central to this effort is the promotion of coordinated vulnerability disclosure as a strategic tool for securing digital infrastructures. This paper examines how the evolving EU regulatory landscape shapes the legal status of vulnerability dis-

closure researchers, highlighting thus the tension between harmonisation efforts between Member States and legal challenges faced by researchers. It explores the interplay between key legislative instruments, including the NIS 2 Directive, the Cyber Resilience Act, and the European Cybersecurity Act, and provides a comparative analysis of national frameworks in France, the Netherlands, and Belgium.

Despite commendable progress in institutionalising cybersecurity obligations across Member States, a significant gap persists in the legal protection afforded to the individuals who initiate vulnerability reporting. The absence of harmonised safeguards for researchers threatens to undermine the effectiveness of EU-wide disclosure mechanisms.

1. The European Legal Framework: Ambitions and Shortcomings

The NIS 2 Directive reflects a marked evolution from its predecessor by explicitly addressing the governance of vulnerabilities. Article 12 requires Member States to ensure that CSIRTs[2] contribute to coordinated vulnerability disclosure. Moreover, Recital 60 acknowledges the precarious legal position of researchers and calls on Member States to provide guidelines aimed at mitigating the risk of criminal and civil liability. However, this provision remains aspirational rather than mandatory, as recitals have no legal binding force. Crucially, the directive does not define key legal terms such as "illegal access" to a computer system, thereby allowing broad discretion at the national level to incriminate such offenses. Article 20, which imposes obligations on essential and important entities to handle vulnerabilities, similarly neglects to define the role of researchers or set minimum standards for their protection.

In parallel, the Cyber Resilience Act introduces mandatory obligations for manufacturers and software vendors to report known vulnerabilities within 24 hours of discovery. Article 11 of the proposed regulation exemplifies a shift toward tighter regulatory control over the product lifecycle, including post-market surveillance. While this approach is necessary to address the growing risks associated with insecure digital products, it implicitly places researchers in a vulnerable position. Their discoveries often trigger these notification duties, yet the legal framework fails to extend them corresponding protections. The regulation thus embodies a structural asymmetry between the obligations of vendors and the exposure of those who enable compliance. It also raises concerns regarding the timelines imposed on organisations, which might result in undue pressure on researchers to withhold or delay disclosure.

The EU Cybersecurity Act mandates ENISA to develop best practices for vulnerability management. It supports the development of voluntary frameworks and standards, notably through engagement with ISO/IEC 29147 and 30111. However, its approach relies on non-binding recommendations, and ENISA has no enforcement authority over Member States. As a result, national

[2] Cyber Security Incident Response Teams.

practices diverge significantly, and many states remain hesitant to embrace legal reform or modify their existing criminal laws. Reports published by ENISA in 2018 and 2022 have repeatedly highlighted the legal uncertainty faced by researchers, yet few Member States have translated these concerns into domestic legislation. Even where ENISA promotes consistent approaches, implementation gaps remain due to the absence of monitoring mechanisms or infringement consequences.

2. Diverging National Practices: Comparative Legal Analysis

2.1 France: Legal Exposure and Centralised Structures

In France, the legal framework for cybersecurity offences remains rooted in the *Loi Godfrain* of 1988. Article 323-1 of the French Penal Code criminalises unauthorised access and persistence in an automated data processing system, irrespective of intent. French jurisprudence, as illustrated by the Bluetouff case, has increasingly adopted a strict interpretation of unauthorised access. In that case, a cybersecurity researcher accessed documents that were publicly available online, without bypassing technical protections. While the trial court acquitted him, considering the access non-fraudulent and the data not stolen in the legal sense, the Court of Appeal and the Court of Cassation upheld his conviction for fraudulent retention in the system and data theft. The case reveals the rigidity of the French criminal framework, which tends to overlook intent or security research motivations in favor of a formalist approach.

This rigidity is compounded by the lack of legislative reform. Although the Law for a Digital Republic (*Loi pour une République numérique*, n° 2016-1321) introduced a framework for rights and transparency, it did not include provisions to (completely) protect ethical hackers. Attempts to introduce a legal exception for good-faith researchers, such as the 2016 "amendement Bluetouff", were ultimately rejected. Article L. 2321-4-1 of the *Code de la défense* imposes reporting obligations on vendors of digital products in cases of "significant vulnerability", yet it remains ambiguous as to the criteria for significance and imposes no parallel obligations on or protections for independent researchers. Despite initial efforts to carve out a form of "right to hack" in cases of immediate disclosure, the French legal framework continues to treat ethical hacking as a punishable offense, providing at best sentence exemptions rather than prosecutorial immunity. This incomplete and deterrent approach stands in contrast with more incentive-based models, such as vulnerability disclosure programs or bug bounty schemes supported by the private sector.

France has centralised vulnerability disclosure through the *Agence nationale de la sécurité des systèmes d'information* (ANSSI), which serves as the primary point of contact. Although ANSSI provides technical guidance and mediates some disclosures, the agency does not possess legal authority to grant immunity or prevent prosecution.

Crucially, while good-faith reporting to ANSSI may reduce the risk of crim-

inal prosecution initiated by the State (according to Article 47 of the *Loi pour une République numérique*), it does not shield the researcher from all possible legal actions. Private entities, such as affected companies or service providers, retain the right to initiate civil or even criminal proceedings against researchers independently. Companies may pursue claims based on alleged breaches of contract, violation of terms of service, unauthorized access under the Law Godfrain or seek damages for alleged harm caused by the disclosure.

Thus, even when researchers act transparently and responsibly by informing ANSSI in good faith, the absence of statutory safe harbor provisions leaves them exposed to legal uncertainties and potential litigation risks originating from the private sector. This gap in protection discourages many security researchers from coming forward, limiting the effectiveness of coordinated vulnerability disclosure mechanisms in France.

Moreover, the increasing volume of incident reports and the expansion of ANSSI's mandate under NIS 2, extending its responsibility to cover more sectors and type of incidents, raise questions about its capacity to engage meaningfully with researchers, particularly those outside institutional settings. Many researchers remain reluctant to come forward in the absence of a predictable and legally secure channel, as stated by ENISA.

2.2 The Netherlands: Pragmatic Legal Innovation

By contrast, the Netherlands has adopted a more pragmatic and decentralised approach. Article 138ab of the Dutch Penal Code criminalises unauthorised access but allows for prosecutorial discretion. Since 2013, the National Cyber Security Centre (NCSC) has published Coordinated Vulnerability Disclosure guidelines that encourage good-faith reporting. These guidelines have no binding legal force, yet their widespread adoption by public and private entities has effectively created a standard. The Public Prosecution Service has issued a formal policy stating that no criminal proceedings will be initiated against researchers who follow the NCSC's guidance. This informal safe harbour is further supported by public statements from the Dutch government, which assure researchers of freedom from civil liability provided that their actions are proportionate, transparent, and responsible.

The Dutch approach is reinforced by jurisprudence, most notably in the case of NXP v. Radboud University. In this 2008 decision, a district court rejected the semiconductor company's attempt to prevent the publication of academic research exposing vulnerabilities in its MIFARE chip. The court held that freedom of scientific expression and the public interest in improved cybersecurity outweighed the commercial interests of the vendor. This case signalled a clear commitment to recognising the societal value of vulnerability research. It also demonstrated that legal protection can emerge through principled judicial reasoning and administrative guidance, even in the absence of formal legislative reform. Importantly, Dutch universities and research centres have internalised this model, leading to institutional CVD policies that support rather than restrict their cybersecurity researchers.

2.3 Belgium: A Hybrid Approach

Belgium offers a hybrid model situated between the centralised French approach and the flexible Dutch regime. Article 550bis of the Belgian Penal Code prohibits unauthorised access to IT systems, yet the application of this provision is tempered in case of researchers who act in good faith, respect confidentiality, and follow CERT.be's disclosure process. The Centre for Cybersecurity Belgium (CCB) formalised its Coordinated Vulnerability Disclosure Policy, establishing clear reporting channels and procedural safeguards. Unlike France, Belgium permits private companies to adopt their own CVD frameworks, with state intervention occurring only in the absence of effective private mechanisms.

This decentralised and dialogue-driven model creates an enabling environment for researchers while preserving state oversight. Though Belgium does not yet codify a full safe harbour into statutory law, the combination of prosecutorial discretion and administrative commitment results in a de facto protective regime. Researchers in Belgium benefit from predictable procedures, documented expectations, and a culture of cooperation fostered by CERT.be's consistent practice of acknowledging contributors.

3. Ongoing Legal Challenges for Researchers

Despite these promising examples, significant obstacles remain at the European level. The absence of a harmonised definition of unauthorised and illegal access to a computer system creates fragmentation. While the Budapest Convention (Articles 2–6) provides a common legal baseline, its interpretation varies widely. French criminal law focuses on the factual absence of authorisation and disregards the researcher's intent. Dutch and Belgian authorities, by contrast, consider the actor's intent and the broader context of their actions. The lack of clarity around the requirement of malicious intent leads to legal uncertainty, especially in jurisdictions that do not distinguish between defensive and offensive behaviour.

Moreover, civil liability continues to be a deterrent. In the absence of explicit immunity provisions, researchers remain vulnerable to lawsuits based on allegations of damage or reputational harm. Even where criminal proceedings are unlikely, the risk of civil litigation persists. France, in particular, has no established jurisprudence or administrative guidance excluding civil claims for proportionate and responsible disclosure. This contrasts with the Netherlands, where government guidance offers assurance against civil reprisals.

The EU Whistleblower Directive also offers limited protection for vulnerabilities researchers. Its scope is restricted to the reporting of breaches of EU law and requires a direct connection between the reported act and an identifiable unlawful activity. Since vulnerability disclosure often concerns potential rather than actual exploitation, the directive's applicability is uncertain. Researchers acting outside an organisational context or without direct evidence of illegality may fall outside its protections.

Legal uncertainty is further exacerbated by the lack of coherent national

strategies. Some Member States continue to ignore ENISA's recommendations or provide only vague mentions of vulnerability disclosure in their cybersecurity roadmaps. This fragmented institutional commitment leads to inconsistent cooperation with researchers, undermining the objective of building an EU-wide culture of security.

4. Recommendations for a Harmonised Legal Framework

To address these deficiencies, a coordinated reform agenda is necessary. First, the introduction of an EU-wide safe harbour provision would provide researchers with conditional immunity from criminal and civil liability, provided they act in accordance with defined principles. These principles should include the absence of malicious intent, the use of secure and confidential reporting channels, the obligation to have such reporting channels, and the provision of adequate time for remediation prior to public disclosure. Such a framework could be integrated into the NIS 2 Directive or appended to the Cyber Resilience Act.

Second, ENISA should be entrusted with the task of developing binding technical standards and procedural templates for CVD, in collaboration with Member States and stakeholders. These standards should be based on international norms such as ISO/IEC 29147 and 30111 but tailored to the legal and institutional diversity of the EU. A centralised platform for disclosure, managed by ENISA or national CSIRTs, could serve as a gateway for protected communication and guarantee consistent treatment across borders.

Third, Member States should include explicit recognition of ethical vulnerability research in their national cybersecurity strategies. This recognition should extend to the provision of legal assistance, certification mechanisms, and participation in public-private initiatives such as bug bounty programs. The example of the "Hack the Pentagon" initiative in the United States illustrates how structured programs can legitimise and channel researcher activity toward national security objectives.

Fourth, the Council of Europe should clarify the application of the Budapest Convention to vulnerability disclosure. An interpretive note should affirm that acts conducted in accordance with recognised CVD frameworks, and in the absence of malicious intent, do not fall within the scope of criminal liability under Articles 2 through 6. This clarification would provide guidance to national courts and reinforce the legitimacy of ethical hacking within the European legal order.

Finally, EU institutions should consider developing a certification scheme for responsible security researchers, linked to European digital identity and trust services. Such a scheme could formalise the status of ethical hackers and facilitate cross-border cooperation, while enhancing legal certainty for all parties involved.

Conclusion

While the European Union has taken important steps toward the institutionalisation of coordinated vulnerability disclosure, the protection of those who enable this process remains inadequate. The current regulatory environment is marked by fragmentation, ambiguity, and uneven enforcement. Researchers face disproportionate risks, and this deters valuable contributions to collective cybersecurity. The experiences of the Netherlands and Belgium demonstrate that coherent legal frameworks, built on trust and mutual recognition, can foster a productive disclosure ecosystem without compromising security or legal certainty. As NIS 2 and the Cyber Resilience Act are implemented, the EU must embed strong safeguards to ensure a secure, ethical, and effective CVD system for Europe's digital future.

References

1 Belgian Law of 28 November 2022 on the protection of persons who report breaches of Union or national law within a legal entity in the private sector. Belgian Official Journal, 2022.

2 Carrapico, H., and Farrand, B. Digital sovereignty and taking back control: From regulatory capitalism to regulatory mercantilism in EU cybersecurity. European Security, 2022.

3 CERT.be. Coordinated vulnerability disclosure policy. Centre for Cybersecurity Belgium, 2021.

4 Council of Europe. Convention on Cybercrime (Budapest Convention), 2001.

5 Court of Appeal of Paris. Judgment of 30 October 2002, Antoine C. v. Tati.

6 Court of Cassation (France). Criminal Chamber, Decision of 20 May 2015, No. 14-81.336 (Bluetouff case).

7 Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law. Official Journal of the European Union, 2019.

8 Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive). Official Journal of the European Union, 2022.

9 ENISA. Developing national vulnerability programmes: Challenges and initiatives. ENISA, 2023.

10 ENISA. Economics of vulnerability disclosure. ENISA, 2018.

11 ENISA. ENISA threat landscape 2022. ENISA, 2022.

12 ENISA. ENISA threat landscape 2023. ENISA, 2023.

13 ENISA. Good practice guide on vulnerability disclosure: From challenges to recommendations. ENISA, 2016.

14 European Commission. Cybersecurity Strategy of the European Union: An open, safe and secure cyberspace. Joint Communication, 7 February 2013.

15 European Commission. Proposal for a regulation on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020 (Cyber Resilience Act). 2022.

16 International Organization for Standardization. ISO/IEC 29147:2018 – Information technology – Security techniques – Vulnerability disclosure. ISO, 2018.

17 International Organization for Standardization. ISO/IEC 30111:2019 – Information technology – Security techniques – Vulnerability handling processes. ISO, 2019.

18 *Loi n○ 88-19 du 5 janvier 1988 relative à la fraude informatique (dite «Loi Godfrain »). Journal Officiel de la République Française*, 1988.âĂÍ (English: French Law No. 88-19 of 5 January 1988 on computer fraud (Godfrain Law), French Official Journal, 1988.)

19 NCSC NL. Coordinated vulnerability disclosure: The guideline. National Cyber Security Centre Netherlands, 2018.

20 NXP Semiconductors Netherlands BV v. Radboud Universiteit Nijmegen. District Court of Arnhem, 2008.

21 OECD. Encouraging vulnerability treatment: Overview for policy makers. OECD Digital Economy Papers, 2021.

22 Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). Official Journal of the European Union, 2019.

23 Schmitz-Berndt, S., & Schiffner, S. (2021). Don't tell them now (or at all) – Responsible disclosure of security incidents under NIS Directive and GDPR. International Review of Law, Computers & Technology, 35(2), 135–150.

24 Vostoupal, J., Stupka, V., Harašta, J., Kasl, F., Loutocký, P., & Malinka, K. (2023). The Legal Aspects of Cybersecurity Vulnerability Disclosure: To the NIS 2 and Beyond (SSRN Working Paper No. 4640775)

VI.8 Online Misogyny and Electoral Campaigns: Digital Violence as a Barrier to Women's Political Participation

Abstract accepted for inclusion in proceedings

Luise Koch (`luise.koch@tum.de`*), joint work with Jürgen Pfeffer, Raji Ghawi, Janina Steinert, TU Munich, Germany*

(An extended version of this work has been submitted to, and is available in, arXiv as a preprint [1].)

Technology-facilitated gender-based violence (TFGBV) has become a growing threat to democracy, particularly during elections. Female candidates worldwide face disproportionate levels of online misogyny, which not only affects their well-being but also has broader implications for democratic representation. This

study investigates how online misogyny against female candidates impacted political engagement during the 2022 Brazilian elections, using a large-scale analysis of 10 million tweets directed at 445 female candidates.

Using a self-trained machine-learning classifier, we detected misogynistic content in Brazilian Portuguese, allowing for a quantitative analysis of how digital harassment affects political discourse. The findings reveal that young, left-wing, and highly visible candidates were significantly more likely to be targeted. Moreover, a higher number of misogynistic attacks in one week correlated with a decline in candidates' online activity in the following week, demonstrating a silencing effect on political participation.

The study highlights the limits of AI-driven content moderation focusing on "toxicity", as misogynistic rhetoric often bypasses detection through subtle linguistic strategies. It also raises questions about platform responsibility, as current regulations and moderation strategies fail to prevent widespread gendered harassment. This research underscores the need for stronger digital governance frameworks that integrate algorithmic transparency, intersectional policy approaches, and platform accountability to ensure safe online spaces for female politicians.

By situating online misogyny within broader debates on digital democracy, election integrity, and platform regulation, this paper contributes to critical discussions on the role of technology in shaping political participation.

References

1 Luise Koch, Raji Ghawi, Jürgen Pfeffer, Janina Isabel Steinert: "Online Misogyny and Electoral Campaigns: Digital Violence as a Barrier to Women's Political Participation", arXiv, submitted on March 2024.
Available at `https://arxiv.org/abs/2403.07523`.