SOBIGDATA++
PLUS PLUS

Deliverable D2.4

Ethics and Legality Framework activities 3

SoBigData Plus Plus receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 871042.

DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData Plus Plus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (60 months)
ENDING DATE	31/12/2024
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042
DELIVERABLE INFORMATION	
WORK PACKAGE	WP2 NA1 - Responsible Data Science
WORK PACKAGE LEADER	TUDelft
WORK PACKAGE PARTICIPANTS	CNR, UNIPI, SSSA, KCL, LUH, CNRS, URV
DELIVERABLE NUMBER and TITLE	D2.4 Ethics and Legality Framework activities 3
AUTHOR(S)	Juan M. Durán (TUDelft)
CONTRIBUTOR(S)	Josep Domingo-Ferrer (URV), Iryna Lishchuk (LUH), Francesca Donati (SSSA), Giovanni Comandé (SSSA), Ilaria Barsanti (CNR)
EDITOR(S)	Valerio Grossi (CNR)
REVIEWER(S)	Ilaria Barsanti (CNR), Marco Braghieri (KCL)
CONTRACTUAL DELIVERY DATE	31/12/2024
ACTUAL DELIVERY DATE	30/12/2024
VERSION	1.1
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	36
KEYWORDS	Privacy, ethics, data science

EXECUTIVE SUMMARY

This deliverable presents a comprehensive report on the activities undertaken, ongoing, and planned between 1st of January 2023, and 31st December 2024, under Work Package 2: NA1 - Responsible Data Science (hereafter WP2) and its associated tasks.

The document is organized as follows: Section 1 outlines the significance of WP2 for other work packages within the SoBigData++ consortium. Section 2 details the general activities conducted by all WP2 members. It is further divided into the following subsections: Subsections 2.1 to 2.4 cover the activities of each individual task as well as collaborative efforts. Subsection 2.5 lists publications by WP2 members. Additionally, the appendices provide supplementary information on the activities described in the report: Appendix A includes statistics on TransNational Access (TNA), as referenced in Subsections 2.1 to 2.4. Appendix B features the white paper to be published by the High-Level Advisory Board.

DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871042.

SoBigData++ strives to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and big data to understand the complexity of our contemporary, globally-interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015. Becoming an advanced community, SoBigData++ will strengthen its tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments.

This document contains information on SoBigData++ core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData++ Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The content of this publication is the sole responsibility of the SoBigData++ Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

GLOSSARY

BOEL	Board for Operational Ethics and Legality
EU	European Union
EC	European Commission
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
RI	Research Infrastructure
VA	Virtual Access
TNA	Transnational Access
XAI	Explainable Artificial Intelligence

TABLE OF CONTENTS

1 Relevance to SoBigData++

This document provides a full report of relevant activities carried out, ongoing, and planned during the period 1st January 2023 - 31st December 2024 by Work Package 2: NA1 - Responsible Data Science (hereby WP2) and its Tasks.

1.1 Relevance to project objectives

This document complies with the objectives established for WP2 consisting in gathering innovative and proactive responses to structural problems currently emerging in social and cultural data analytics, such as online information disorder, the Facebook data privacy breach and algorithmic bias and discrimination. This is ensuring that the project not only develops best practices and resources for social and cultural data analytics practitioners, and it is granting a wider, informed, engaged and equitable participation and impact. To this end, the tasks related to WP2 have carried out a series of activities as described in section 2.

The document corresponds to deliverable D2.4: Ethics and Legality Framework activities 3 according to WP2. The deliverable must be a report describing activities performed in the WP2 as a whole and specifically activities carried out in the various boards and tasks.

1.2 Relation to other work packages

WP2 carried out and planned different activities independently as well as in close collaboration with several other WPs. These activities focus on emerging ethical and social concerns that transpire from the analysis and validation of usage of data mining resources. Of particular relevance are the following WPs:

- WP3: Dissemination, Impact, and Sustainability
- WP4: Training
- WP5: Accelerating Innovation
- WP6: Transnational Access
- WP7: Virtual Access
- WP8: Social Mining and Big Data Resource Integration
- WP10: Exploratories

Naturally, the BOEL as well as the High-Level Advisory Board is attending all ethical and legal consultations required within SoBigData++ by any WP and/or its members.

1.3 Structure of the document

The document is structured as follows.

- Section 2 reports on the general activities carried out by all the members of WP2. Section 2 is subdivided into the following sections: Subsections 2.1 to 2.4 report on each individual task's activities as well as collaborations; subsection 2.5 reports on the publications by the members of WP2.

- Finally, the appendixes include extra details about activities mentioned in this deliverable. Appendix A reports on the statistics of TNA (as indicated in Sections 2.1 to 2.4) and appendix B contains the white paper 2024 outlining the consortium's ethical and political position on Big AI to be published by the High-Level Advisory Board.

2 Report on WP2 activities

Activities carried out by WP2 partners under SoBigData++ acknowledgment:

URV: In the timeframe between 1st of January 2023 and 31st of December 2024, URV has organized the following activities:

Invited Talks:

- Josep Domingo-Ferrer, ``The role of AI in education'', invited talk, Notre Dame University, Indiana, USA, Dec. 1, 2023.
- Josep Domingo-Ferrer, ``Privacy and security in centralized and decentralized machine learning'', invited talk, King Abdullah University of Science and Technology, Jeddah, Saudi Arabia, Nov. 9, 2023.
- Josep Domingo-Ferrer, ``Què és i què pot fer la intelligència artificial?'' (in Catalan, ``What is AI and what can it do?''), invited talk at the Section of Philosophy and Social Sciences of Institut d'Estudis Catalans, Barcelona, Catalonia, Oct. 19, 2023.
- Josep Domingo-Ferrer, ``On the Use (and Misuse) of Differential Privacy in Machine Learning'', keynote talk at EPIA 2023-22nd Portugueses Conference on Artificial Intelligence, Horta, Faial Island, Açores, Portugal, Sep. 6, 2023.
- Josep Domingo-Ferrer and Alberto Blanco-Justicia, ``Privacy-preserving techniques and approaches (+ Hands-on session)'' invited talks, SoBigData Summer School 2023, Lipari, Eolian Islands, Italy, July 11, 2023.
- Josep Domingo-Ferrer, ``Using co-utility to achieve security and privacy in federated learning and fully decentralized learning'', invited talk at IAIL 2023- Imagining the AI Landscape after the AI Act, Munich, Germany, June 27, 2023.
- Josep Domingo-Ferrer, ``Dades sintètiques i privadesa'' (``Synthetic data and privacy'', in Catalan), invited talk, Catalan Data Protection Agency, Mar. 15, 2023. Given at Catalonia's School of Public Administration.

Event organized:

- In 2023, we have started the preparation of the "PSD 2024-Privacy in Statistical Databases" conference, to be held on September 25-27, 2024. SoBigData++ is among the conference sponsors: https://urldefense.com/v3/__https://unescoprivacychair.urv.cat/psd2024/sponsors__;!!PAKc-5URQlI!8RYUZ_ypclC7TaDu1lmkG8AyO3tPHouN8ZFd6AeZVnoL5FnNcCyss-LojAG6duwn6vl6UsxY_D6Rf5q_iStrqwcW-5gZ$ https://crises-deim.urv.cat/psd2024/sponsors

CNR: In the timeframe between 1st of January 2023 and 31st of December 2024, CNR has organized the following activities:

Events Organization:

- Organization of the SoBigData Summer School "Responsible Data Science for Society: Models, Algorithms, Trustworthy AI", held in Lipari on July 9-15, 2023.
- Organization of the SoBigData Summer School "Responsible Data Science for Society: Models, Algorithms, Trustworthy AI", held in Baratti (LI), Italy on June 16-22, 2023.
- Organization of the 2nd Workshop "IAIL 2023: Imagining the AI Landscape after the AI Act", held in Munich on June 27 and co-located with the Hybrid Human-Artificial Intelligence - HHAI 2023.

- Organization of the event Empowering Enterprises through BigData and AI The European Data Strategy 2024, Bruxelles, February 13, 2024
- Organization of the conference Discovery Science 2024, Pisa, 14-16 October 2024.

Events Participation:

- Participation as speaker in the "Digital with Purpose Global Summit 2023", held in Lisbon on September 27 (panel title "Big Data and health care" by Roberto Trasarti (ISTI-CNR))
- Participation as speaker in the "Raising Awareness on Data Altruism between Gaps and Enablers" online workshop on December 6 (talk title: "Ongoing initiatives on data altruism in specific sectors: The SoBigData Infrastructure" by Francesca Pratesi (ISTI-CNR)).
- Participation as speaker in the "AI-GAP 2023: Algorithmic Biases in Artificial Intelligence from Interdisciplinary Perspectives" workshop, held in L'Aquila on November 20 (talk title: "Empowering Change: SoBigData RI's Initiatives for Gender Inclusion" by Michela Natilli (ISTI-CNR)).
- Participation as speakers in the "ITADATA 2023: Italian Conference on Big Data and Data Science", held in Naples on September 13 (tutorial title: "Open the black box: Methods and Practices for explainable AI" by Michela Natilli (ISTI-CNR) and Carlo Metta (ISTI-CNR)")
- Participation as speakers in the SoBigData Summer School, with the following presentations/tutorial (highlighted the CNR people):
    - The SoBigData RI, Roberto Trasarti (ISTI-CNR)
    - Privacy Risk Assessment and Vulnerabilities: from theory to practice, Josep Domingo Ferrer, Alberto Blanco Justicia, Roberto Pellungrini, Francesca Pratesi (ISTI-CNR)
    - Social Artificial Intelligence, Dino Pedreschi, Salvatore Rinzivillo (ISTI-CNR)
    - Explainable Machine Learning for Trustworthy AI, Fosca Giannotti, Andrea Beretta (ISTI-CNR), Anna Monreale
    - Explainability Techniques for Tabular Data, Images, Time Series and Graphs, Anna Monreale, Carlo Metta (ISTI-CNR), Riccardo Guidotti, Giovanni Stilo (ISTI-CNR), Mario Alfonso Prado Romero, Francesca Naretto
    - Exploratories Research Highlight, Luca Pappalardo (ISTI-CNR), Donia Kamel, Todor Galev, Angelo Facchini, Carolina Scarton

The role of Francesca Donati in WP2 was as an external consultant is to evaluate the project proposals submitted within SoBigData++ from an ethical and legal point of view. In particular – given Donati's background – her activity is focused on evaluating the proposals mainly from the point of view of their compliance with the EU Regulation 2016/679 GDPR. The aim is not only to evaluate projects but also to promote among young researchers an ethics for the processing of personal data so that people's rights are not violated. WP2 met in virtual mode almost once a month. Donati was able to attend all the meetings. During each of the meetings, WP2 took stock of the activities carried out and planned subsequent activities. In there, we discussed about how to promote consciousness among the researchers on how to correctly process personal data.

In fact, we noticed that many applicants did not focus on the legal issues that processing personal data implies, for example the origin of the dataset, issues about intellectual property and the lack of consent of the data subjects. During this period, the group discussed the need to request more information on the proposed projects regarding the datasets used. Therefore, changes were made to the project proposal forms to obtain more information from applicants.

During the year 2023, the number of applications has increased, thanks also to the end of the pandemic restrictions. We have noticed that researchers are paying more attention to legal and ethical issues that may arise in their research projects, thanks in part to our guidance. Our role has not been very easy, as GDPR does not apply worldwide, and many applicants are also extra UE.

TUDelft: In the timeframe between 1st of January 2023 and 31st of December 2024, TU Delft carried out the following activities:

- Participation as Keynote Speaker: "Responsible Data Science: what is it and why do we want it?". Summer School "Responsible data science for society: Models, Algorithms, Trustworthy AI" - SoBigData++, Juan M. Durán.
- Science Cafe "Justification in forensic machine learning: what is it and why do we want it?" https://www.tudelft.nl/en/events/2023/safety-and-security-institute/tu-delft-nfi-science-cafe-deepdive-edition, Juan M. Durán
- Book edition "Philosophy of science meets Machine Learning - Core issues, new perspectives". Forthcoming 2025. Juan M. Durán and Giorgia Pozzi (eds)
- Invited talk: "Models, Representation, and Computation", celebration of Morrison and Humphreys's legacy." Université Paris. Juan M. Durán. 2024
- Invited talk: "Epistemic opacity, transparency, and computational reliabilism", Ethics and Epistemology of AI reading group. University of Twente. Juan M. Durán. 2024
- Invited talk: "Toward an epistemology for algorithms: computational reliabilism and its limits." Theoretical Philosophy Colloquium. Utrecht University. Juan M. Durán. 2024
- Invited talk: "Justification, transparency, and computational reliabilism." Seminar at the Computer Science Department. Utrecht University. Juan M. Durán. 2023
- Invited talk: "Sources of bias in AI for justice". Shaping the Future of Forensics with Proteomics. TU Delft, UCLan & Forens-OMICS, Erasmus MC, UGent, UVA. Juan M. Durán. 2023
- Invited talk: "Computational Reliabilism, justification, and Machine learning." Dalle Molle Institute for Artificial Intelligence. Università della Svizzera italiana. Juan M. Durán. 2023
- Invited talk: ""Justification and ML: What Computational Reliabilism tell us." HLRS - University of Stuttgart." Juan M. Durán.
- Outreach: "SoBigData Digital Coffee: a shot of Big Data into Innovation, Entrepreneurship, Ethics and Impact. 3°webinar - Webinar 3 — Responsible Innovation: Data Ethics in business and society." Juan M. Durán. 2024
- European Parliament: "Empowering Enterprises through BigData and AI. The European Data Strategy 2024" http://datastrategy2024.isti.cnr.it/index.html. Juan M. Durán, 2024

2.1 Task 2.1. Board of Operational Ethics and Legality

Task leader: TUDelft
Participants: LUH, CNR, SSSA

The Board of Operational Ethics and Legality (BOEL) oversees the Micro-Project "BOEL Works," which offers a recommendation service exclusively designed to address legal, ethical, and societal concerns arising from the research conducted by members of SoBigData++. This service provides guidance on methodological approaches—including references, tools, standards, and policies—applicable in various research contexts, such as grant applications, academic publications, databases, research outreach, and workshops. The goal is to assist members in effectively managing ethical, legal, and societal dimensions of their work.

Through this initiative, BOEL fosters heightened individual and collective awareness of the ethical, regulatory, and societal implications of data science, while promoting accountable and informed problem-solving in research endeavors.

The "BOEL Works" Micro-Project is an ongoing effort, renewed biannually. This report provides an overview of the services rendered by BOEL in response to requests received between 1st January 2023, and 31st December 2024. During this period, most requests pertained to TNA. Instances of non-TNA requests are explicitly noted. For clarity, the activities of BOEL within the specified timeframe are presented according to the distinct phases of the Micro-Project's implementation.

2.2 Task 2.2 Bottom-up Ethics and Legality for Data Science

Task leader: SSSA
Participants: TUDelft, CNR, URV, KCL, SSSA, LUH

The SSSA team (Giovanni Comandè, Magali Contardi) has actively taken part in the monthly WP2 meetings as well as the general consortium meetings.

From a scientific viewpoint, SSSA contributed to the debate on AI, big data, ethics and regulation with scientific publications (see section 2.5).

2.3 Task 2.3 High-Level Advisory Board

Task leader: LUH
Participants: TUDelft, CNR, UNIPI, URV, CNRS, SSSA

2.3.1 Activities performed

Another activity performed by WP2 within Task 2.3 was the production of white papers. The white paper is entitled "Democracy in the digital age" and was authored by Jeroen Van den Hoven, Giorgia Pozzi, Marc Stauch, Francesca Musiani, Josep Domingo-Ferrer, Salvatore Ruggieri, Francesca Pratesi, Roberto Trasarti and Giovanni Comandé. WP2 completed the white paper and submitted it for publication but it was unfortunately rejected. Thus, WP2 has decided to make the article publicly available.

Likewise, after some initial delays, we have now published the second paper, "Artificial Intelligence across Geo-political Models of Digital Governance," in the LUH repository (see https://www.repo.uni-hannover.de/handle/123456789/18086 and the English version of the e-mail below). The paper is currently licensed under a cc by license for non-commercial use. The DOI is already created in the document and should be available soon.

2.3.2 Participants involved

The High-Level Advisory Board consists of the following members:

- TUDelft: Jeroen van den Hoven (Chair)
- LUH: Marc Stauch (Vice Chair)
- SSSA: Giovanni Comandé (Vice Chair)
- CNR: Fosca Giannotti (regular member)
- UNIPI: Salvatore Ruggieri (regular member)
- URV: Josep Domingo-Ferrer (regular member)
- CNRS: Francesca Musiani (regular member)
- Giovanni Sartor – UEI (external expert)

2.3.3 Goals of the activities

The annual white papers have been produced and disseminated through various channels (T.3.2, T.3.4, and T.5.1) to influence European society at multiple levels. These objectives have been successfully achieved in alignment with the agreement. In 2023/24 there took six Board Meetings place and different small meetings in between. find the dates below:

- 23 February 2023
- 29 March 2023
- 12 September 2023
- 8 November 2023
- 14 December 2023
- 22 February 2024.

Furthermore, the board participated in the review of Transnational Access Requests, specifically assessing the data protection and privacy aspects of the requests.

2.3.4 Outcome(s) produced

- "Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications to Consider" Jeroen van den Hoven, Giovanni Comandé, Salvatore Ruggieri, Josep Domingo-Ferrer, Francesca Musiani, Fosca Giannotti, Marc Stauch, and Iryna Lishchuk. https://www.opiniojurisincomparatione.org/articles/towards-a-digital-ecosystem-of-trust-ethical-legal-and-societal-implications/
- "Democracy in the Digital Age" (2024), Jeroen Van den Hoven, Giorgia Pozzi, Marc Stauch, Francesca Musiani, Josep Domingo-Ferrer, Salvatore Ruggieri, Francesca Pratesi, Roberto Trasarti and Giovanni Comande at: https://www.repo.uni-hannover.de/
- "The European Approach to Artificial Intelligence across Geo-political Models of Digital Governance" (2024), Jeroen Van den Hoven, Giorgia Pozzi, Marc Stauch, Iryna Lishchuk, Francesca Musiani, Josep Domingo-Ferrer, Salvatore Ruggieri, Francesca Pratesi, Roberto Trasarti and Giovanni Comande at: https://www.repo.uni-hannover.de/handle/123456789/18086

With the completion of these two white papers, WP2 has fulfilled its commitment to publish and disseminate findings through various channels.

2.3.5 Contribution to the task and WP2 in general

Annual reports are prepared to highlight best practices, emerging trends, and innovative approaches developed within the project, along with providing guidance and forming high-level perspectives on the legal and ethical issues arising from project activities. All members of the High-Level Advisory Board (HLAB) are actively involved in discussions and in drafting the second and third white papers, which focus on the activities and strategic positioning of SoBigData++.

2.4 Task 2.4 Critical Data Literacy

Task leader: KLC
Participants: LUH, TUDelft, CNR, SSSA

The KLC team (Mark Coté) has actively taken part in the monthly WP2 meetings as well as the general consortium meetings. From a scientific viewpoint, KLC contributed to the debate on AI, big data, ethics and regulation with scientific publications (see Section 2.5).

KLC has also continue its collaboration with CNR and the project management team to identify all Open-Access articles that were published before until the 31st of December 2024 that also had acknowledged SoBigData++ in the correct manner [i.e., "acknowledge support from EU HORIZON 2020 INFRAIA-2019-1(SoBigData++) No. 871042"][1].

A test-page was created for a sample article, which can be found here:

https://sobigdata.d4science.org/group/sobigdataliteracy/literature?path=/dataset/ba441f5d-0260-43d2-b1cd-c418ff2c01a3

Once it was uploaded, WP3 and the project management team both agreed that a new item page should be developed, providing further emphasis on the authors of the paper, its green access status and eliminating some of the visible fields that in the item creation interface process are currently compulsory. Further activity was outlined, planning an interaction with WP7 (Virtual Access) and WP9 (E-Infrastructure and Supercomputing Network) to develop an ad-hoc item creation page for Data Literacy.

2.5 Publications by members of WP2

1. Josep Domingo-Ferrer and Melek Önen (eds.), Privacy in Statistical Databases-PSD 2024, Lecture Notes in Computer Science, vol. 14915, Springer, 2024. ISBN 978-3-031-69650-3
2. Faisal Ahmed, David Sánchez, Zouhair Haddi, and Josep Domingo-Ferrer, "MemberShield: a framework for federated learning with membership privacy", Neural Networks, to appear.
3. Younas Khan, David Sánchez, and Josep Domingo-Ferrer, "Federated learning-based natural language processing: a systematic literature review", Artificial Intelligence Review, to appear.

[1] All files were downloaded from their hosting platforms, renamed, and uploaded in a dedicated folder in the RI. at the following folder: https://data.d4science.net/WXLZ

4. Liangyu Zhong, Lulu Wang, Lei Zhang, Josep Domingo-Ferrer, Lin Xu, Changti Wu, and Rui Zhang, “Dual-server based lightweight privacy- preserving federated learning”, IEEE Transactions on Network and Service Management, vol. 21, no. 4, pp. 4787-4800, 2024.
5. Krishnamurty Muralidhar, Steven Ruggles, Josep Domingo-Ferrer, and David Sánchez, “The counterfactual framework in Jarmin et al. is not a measure of disclosure risk of respondents”, PNAS-Proceedings of the National Academy of Sciences, 121(11) e2319484121, 2024. https: //doi.org/10.1073/pnas.2319484121.
6. Najeeb Moharram Jebreel, Josep Domingo-Ferrer, David Sánchez, and Alberto Blanco-Justicia, “LFighter: Defending against the label-flipping attack in federated learning”, Neural Networks, vol. 174, pp. 111-126, 2024.
7. Najeeb Jebreel, Josep Domingo-Ferrer, Alberto Blanco-Justicia, and David Sánchez, “Enhanced security and privacy via fragmented federated learning”, IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 5, pp. 6703-6717, 2024.
8. Krishnamurty Muralidhar and Josep Domingo-Ferrer, “A Rejoinder to Garfinkel (2023) — Legacy statistical disclosure limitation techniques for protecting 2020 Decennial US Census: still a viable option”, Journal of Official Statistics, vol. 39, no. 3, pp. 411-420, 2023.
9. Jesús Manjón, Josep Domingo-Ferrer, David Sánchez, and Alberto Blanco- Justicia, “Secure, accurate and privacy-aware fully decentralized learning via co-utility”. Computer Communications, vol. 207, pp. 1-18, Jul. 2023.
10. David Sánchez, Josep Domingo-Ferrer, and Krishnamurty Muralidhar, “Confidence-ranked reconstruction of census records from aggregate statistics fails to capture privacy risks and re-identifiability”, PNAS- Proceedings of the National Academy of Sciences, 120 (18) e2303890120, Apr. 24, 2023. https://doi.org/10.1073/pnas.2303890120
11. Krishnamurty Muralidhar and Josep Domingo-Ferrer, “Database reconstruction is not so easy and is different from reidentification”, Journal of Official Statistics, vol. 39, no. 3, pp. 381-398, 2023.
12. Ashneet Khandpur Singh, Alberto Blanco-Justicia, and Josep Domingo- Ferrer, “Fair detection of poisoning attacks in federated learning on non- i.i.d. data”, Data Mining and Knowledge Discovery, vol. 37, no. 5, pp. 1998-2023, 2023.
13. Najeeb Jebreel and Josep Domingo-Ferrer, “FL-Defender: combating targeted attacks in federated learning”, Knowledge-Based Systems, vol. 260, p. 110178, 2023.
14. Alberto Blanco-Justicia, David Sánchez, Josep Domingo-Ferrer and Krishnamurty Muralidhar, “A critical review on the use (and misuse) of differential privacy in machine learning”, ACM Computing Surveys, vol. 55, no. 8, pp. 1-16, 2023.
15. Rami Haffar, David Sánchez, and Josep Domingo-Ferrer, “Explaining predictions and attacks in federated learning via random forests”, Applied Intelligence, vol. 53, no. 1, pp. 169-185, 2023.
16. Josep Domingo-Ferrer and Jesús Manjón, “Circuit-free general-purpose multi-party computation via co-utile unlinkable outsourcing”, IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 1, pp. 539-550, 2023.
17. Fadi Hassan, David Sánchez and Josep Domingo-Ferrer, “Utility- preserving privacy protection of textual documents via word embeddings”, IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 1058-1071, 2023.
18. David Sánchez, Najeeb Jebreel, Krishnamurty Muralidhar, Josep Domingo-Ferrer, and Alberto Blanco Justicia, “An examination of the alleged privacy threats of confidence-ranked reconstruction of Census microdata”, in Lecture Notes in Artificial Intelligence, vol. 14915, pp. 213-224. Vol. Privacy in Statistical Databases (PSD 2024), Antibes Juan- les-Pins, France, Sep. 25-27, 2024.
19. Rami Haffar, Francesca Naretto, Anna Monreale, David Sánchez, and Josep Domingo-Ferrer, “GLOR-FLEX: Local to global rule-based explanations for federated learning”, 2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Yokoahama, Japan, June 30-July 5, 2024.
20. Najeeb Jebreel, Josep Domingo-Ferrer, and Yiming Li, “Defending against backdoor attacks by layer-wise feature analysis (extended abstract)”, in Proc. of the International Joint Conference on Artificial Intelligence (IJCAI 2024), Jeju, Korea, pp. 8416-8420, 2024.

21. Najeeb Jebreel, Josep Domingo-Ferrer, and Yiming Li, “Defending against backdoor attacks by layer-wise feature analysis”, in Lecture Notes in Artificial Intelligence, vol. 13936, pp. 428-440. Vol. The 26th Pacific- Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2023), Osaka, Japan, May 25-28, 2023. Best paper award. CORE A. 869 submissions, 143 accepted papers (16.5% acceptance rate).
22. Paper “The need for a new “right to refuse” the results of emotion recognition AI” by R. Savella, accepted at the workshop “IMAGINING THE AI LANDSCAPE AFTER THE AI ACT” at the International Conference “Hybrid Human Artificial Intelligence (HHAI) 2024” (Malmo, 10-14 June 2024) – pending publication on CEUR Workshop Proceedings (CEUR-WS.org).
23. Paper “FRIA implementation model according to article 27 AI Act” by L. Gatt, M. C. Gaeta, I. A. Caggiano, L. Aulino, E. Troisi, L. Izzo, R. Savella, R. Trasarti, F. Pratesi, accepted for publication at the IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering - IEEE MetroXRAINE 2024, pending publication in the conference proceedings.
24. Abstract "Addressing the environmental impact of AI systems: opportunities and challenges of the AI Act" by R. Savella, accepted at the 1st Doctoral Workshop on Law, Society and Artificial Intelligence that will be held in Pisa on the 2nd of December 2024.
25. FASE – 26th International Conference on Fundamental Approaches to Software Engineering. Paris, France. https://etaps.org/2023/conferences/~40 Giordano d'Aloisio - Giordano D'Aloisio, Antinisca Di Marco and Giovanni Stilo. Democratizing Quality-Based Machine Learning Development through Extended Feature Models - 22-27 April 2023.
26. WSDM - Sixteenth ACM International Conference on Web Search and Data Mining. Singapore. https://www.wsdm-conference.org/2023/ >~200 Mario Alfonso Prado-Romero Mario Alfonso Prado-Romero, Bardh Prenkaj, and Giovanni Stilo. 27 February 2023.
27. Perfail 2024 Third International Workshop on Negative Results in Pervasive Computing. Biarritz, France https://perfail-workshop.github.io/2024/ Andrea d'Angelo Echocardiographic Epicardial Adipose Tissue Quantification: Challenges and Insights. Payel Patra, Andrea Bianchi, Daniele Di Pompeo, Antinisca Di Marco. March 2024
28. Giordano d’Aloisio, Andrea D’Angelo, Antinisca Di Marco, Giovanni Stilo, Debiaser for Multiple Variables to enhance fairness in classification tasks, Information Processing & Management, Volume 60, Issue 2, 2023, 103226, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2022.103226. (https://www.sciencedirect.com/science/article/pii/S0306457322003272)
29. Korte, Data protection in eSports – the processing of performance data to optimise gaming behaviour in Seckelmann/Woerlein, eSport in Law and Society, 2023, 143-166
30. Korte, Cloud-based digital health applications in ZD-Aktuell 2023, 01465
31. Korte, National competences for the AI Regulation in ZD-Aktuell 2024, 01773
32. Korte, GDPR Company-related fines – all clear! Or not? In ZD-Aktuell 2024, 01500
33. Schlee, ChatGPT – Data protection officers have questions and demand answers in ZD-Aktuell 2023, 01204.
34. Schlee, Automated credit assessment = automated individual case decision (?) in ZD-Aktuell 2023, 01161.
35. Fosca Giannotti, Riccardo Guidotti, Anna Monreale, Luca Pappalardo, Dino Pedreschi, Roberto Pellungrini, Francesca Pratesi, Salvatore Rinzivillo, Salvatore Ruggieri, Mattia Setzu and Rosaria Deluca, Trustworthy AI at KDD Lab, Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops, May 29-30, 2023, Pisa, Italy, CEUR workshop proceedings - volume 3486
36. Desara Dushi, Francesca Naretto, Francesca Pratesi, Imagining the AI Landscape after the AI Act, Proceeding of the HHAI 2023 conference, June 26-27, 2023, Munich, Germany, CEUR workshop proceedings - volume 3456

Papers in preparation:

We start with the HLAB paper whose preparation started in 2023. We then mention some preprints on ethics-by-design technologies that we posted in arxiv during 2023.

- J. van den Hoven, G. Pozzi, M. Stauch, F. Musiani, J. Domingo-Ferrer, S. Ruggieri, F. Pratesi, R. Trasarti, and G. Comandè, "Democracy in the digital age", HLAB-WP2 White Paper, in preparation.
- David Sánchez, Najeeb Jebreel, Josep Domingo-Ferrer, Krishnamurty Muralidhar, Alberto Blanco-Justicia: An Examination of the Alleged Privacy Threats of Confidence-Ranked Reconstruction of Census Microdata. CoRR abs/2311.03171 (2023)
- Rami Haffar, David Sánchez, Josep Domingo-Ferrer: Multi-Task Faces (MTF) Data Set: A Legally and Ethically Compliant Collection of Face Images for Various Classification Tasks. CoRR abs/2311.11882 (2023)
- Jordi Soria-Comas, David Sánchez, Josep Domingo-Ferrer, Sergio Martínez, Luis Del Vasto-Terrientes: Conciliating Privacy and Utility in Data Releases via Individual Differential Privacy and Microaggregation. CoRR abs/2312.13712 (2023)

3 Conclusions

This document provides a comprehensive overview of the scope, activities, and team members involved in recent, ongoing, and planned initiatives under Work Package 2: NA1 - Responsible Data Science. The reporting period spans from January 1, 2023, to December 31, 2024.

It highlighted the contributions of Task leaders and participants, detailing their efforts in dissemination through academic conferences, public outreach activities, and publications. These include academic journal articles and broader audience-targeted pieces, many of which are openly accessible.

A key achievement of this Work Package is the creation of a White Paper outlining the consortium's ethical and political position on Big AI, as represented by the members of SoBigData++. The document also contains a draft version of this White Paper 2024.

Appendix A. TransNational Access – Statistics

A link to every project is available upon request.

Submission Date	Type	Answer date
2023-02-03	TNA project	2023-02-13
2023-02-13	TNA integration	2023-02-13
2023-03-13	TNA project	2023-03-27
2023-03-16	TNA project	2023-03-27
2023-03-16	TNA project	2023-04-24
2023-03-13	TNA integration	2023-05-12
2023-06-06	TNA project	2023-06-24
2023-06-12	TNA project	2023-06-20
2023-06-22	TNA integration	2023-06-24
2023-06-26	TNA integration	2023-07-17
2023-06-26	TNA integration	2023-07-17
2023-07-31	TNA project	2023-08-03
2023-08-07	TNA project	2023-08-12
2023-09-19	TNA project	2023-09-29
2023-10-09	TNA project	2023-11-02
2023-10-19	TNA project	2023-10-31
2023-10-25	TNA integration	2023-11-02
2023-11-02	TNA project	2023-11-10
2023-11-06	TNA integration	2023-11-27
2023-11-16	Research project	2023-11-21
2023-11-27	TNA project	2023-11-30
2023-11-27	TNA project	2023-11-30
2023-11-30	TNA integration	2023-12-19
2023-12-04	TNA project	2023-12-19
2023-12-04	TNA project	2023-12-19
2023-12-07	TNA project	2023-12-19
2024-01-04	TNA integration	2024-01-25
2024-01-04	TNA project	2024-02-06
2024-02-01	TNA project	2024-02-08
2024-02-05	TNA project	2024-02-06

2024-02-05	TNA project	2024-02-08
2024-02-06	TNA project	2024-02-15
2024-02-08	TNA integration	2024-02-08
2024-02-13	TNA integration	2024-02-14
2024-02-20	TNA project	2024-03-04
2024-02-22	TNA integration	2024-03-04
2024-03-11	TNA project	2024-03-12
2024-03-11	TNA project	2024-03-29
2024-03-21	TNA project	2024-03-29
2024-03-25	TNA project	2024-04-18
2024-04-02	TNA integration	2024-05-17
2024-04-08	TNA project	2024-04-26
2024-04-08	TNA integration	2024-05-17
2024-04-11	TNA project	2024-04-26
2024-04-26	TNA integration	2024-05-03
2024-05-09	TNA integration	2024-05-25
2024-05-20	TNA project	2024-06-08
2024-05-21	TNA project	2024-06-08
2024-05-23	TNA integration	2024-06-01
2024-05-24	TNA project	2024-06-15
2024-06-04	TNA project	2024-06-25
2024-06-04	TNA project	2024-06-26
2024-06-06	TNA project	2024-06-28
2024-06-18	TNA project	2024-07-10
2024-07-02	TNA project	2024-07-24
2024-07-02	TNA project	2024-07-24
2024-07-11	TNA project	2024-08-02
2024-07-22	TNA project	2024-08-13
2024-08-22	TNA project	2024-09-13
2024-08-29	TNA project	2024-09-11
2024-09-09	TNA project	2024-09-21
2024-10-10	TNA project	2024-10-19
2024-10-15	TNA project	2024-10-22

Appendix B. White paper 2024

Democracy in the Digital Age

Jeroen van den Hoven and Giorgia Pozzi,[1]
Marc Stauch,[2]
Francesca Musiani,[3]
Josep Domingo-Ferrer,[4]
Salvatore Ruggieri,[5]
Francesca Pratesi and Roberto Trasarti,[6]
Giovanni Comandé[7]

[1] Delft University of Technology, Jaffalaan 5, 2628BX, Delft, The Netherlands
[2] Gottfried Wilhelm Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany
[3] Centre Internet et Société, CNRS - Centre national de la recherche scientifique, 59-61 rue Pouchet, 75017 Paris, France
[4] Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalonia
[5] Università di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy
[6] National Research Council of Italy (CNR), Via Giuseppe Moruzzi, 1, 56124 Pisa, Italy
[7] Scuola Superiore Sant'Anna di Pisa, Piazza Martiri della Libertà, 33, 56127 Pisa, Italy

Abstract. Democracy has been greatly affected by the digital age, and this impact will only increase as technology continues to advance at an unprecedented rate. The digital era has transformed the landscape of democratic structures, presenting both opportunities and challenges for the preservation and evolution of democratic principles. As the world becomes more connected through the use of digital technologies, it is important to understand how this affects the democratic process. The digital age has both positive and negative effects on democracy. On the one hand, digital technologies have made it easier for citizens to participate in democracy by facilitating contact with elected officials and increasing opportunities for political engagement through social media and other online platforms. On the other hand, the digital age has also brought challenges to democracy, such as the spread of misinformation and fake news on social media and online platforms, as well as concerns about the security of online voting and potential interference in electoral processes by foreign actors. In addition, the increasing commercialisation of online spaces and algorithms designed to personalise content also pose a threat to democracy by reducing people's exposure to diverse viewpoints and reinforcing pre-existing biases. In the face of these challenges, it is crucial to find ways to address the negative impacts of digital technologies on democracy, while harnessing their positive potential for increased citizen engagement and democratic inclusiveness.

It is important to note that possible measures need to be continuously evaluated and adapted as technology advances in order to ensure the resilience of democratic processes in the face of new challenges posed by the digital age. In conclusion, the digital age offers both opportunities and challenges for democracy. As technology continues to evolve at an unprecedented pace, it is crucial for governments and citizens alike to remain vigilant in monitoring the impact of digital technologies on democracy and to take measures to mitigate any negative effects that arise. In particular, the integrity of the judiciary, freedom of the press, and the protection of free speech face unprecedented challenges in the digital realm. Judicial independence may be undermined by the pressure of public opinion amplified through online platforms, while the erosion of journalistic standards due to the spread of disinformation compromises the role of a free press as a democratic pillar. Simultaneously, freedom of speech is complicated by the tension between open discourse and the need to combat harmful content and manipulation.

Overall, it can be argued that the digital age has had a significant impact on democracy and has necessitated a reassessment of how to ensure its continued success in the 21st century. It is therefore important to develop policies that address both the positive and negative aspects of the use of technology for democracy. To safeguard democratic resilience, a multidisciplinary approach is required, engaging governments, civil society, and the private sector in the development of policies that address these emerging threats.

Keywords: Digital Governance, Digital Democracy

Introduction

The advent of the digital age has brought about a significant transformation in democratic structures, thereby necessitating an in-depth examination of the opportunities and challenges that this transformation presents.

Those who are willing and capable to invest in social media strategies, engage in micro-targeting, psychological profiling and psychography, nudging, machine learning, those willing to mobilize bot nets and launch large volumes of fake Facebook and Twitter accounts, sending AI powered messages, while using Big Data collated from user generated data and massive A/B testing, those deploying recommender systems, managing Youtube Channels: those people can achieve a lot. The bad news is that these things are for sale. it is for sale, as the Cambridge Analytica case has demonstrated. New investigations of foreign money streams and manipulative practices associated with the Brexit campaign in the UK are under investigation. Advanced behavioral science (choice modelling, study of cognitive biases, nudging techniques) machine learning and Big Data can undermine human autonomy and self-determination. The return on investment of learning about the probabilities of winners and losers in elections or referenda can be high. Being in a position to determine the outcomes of these processes with near certainty is of course priceless.

Political investors can exploit the fact that we have introduced vulnerabilities into western liberal democracies and added to the vulnerabilities that were associated with democracy long before the arrival of Information technology. We have naively assumed over the last three decades that the invisible digital hand would by itself produce a good and democratic digital society. This turns out to have been a fatal mistake. Another fundamental mistake has been that the initial promises about the tremendous democratic potential of the Internet has never been realized.

This has been a motor of our contemporary crisis in democratic systems. The crisis is one of a jeopardized pursuit of truth. Our former truth tracking institutions are no longer fully truth oriented. Merchants of doubt have tainted science with anomalies and flawed statistics, influencers and meddlers have compromised independent journalism with fake news, propaganda, disinformation and false hyper-partisan narratives. Lobbyist have undermined trust in politics. Profit maximizing managing elites have obliterated trust in the financial sector and the corporate world. And digital technologies have given a helping hand to all of them.

The Top Legal advisor of the European Commission Paul Nemitz discusses the enormous concentration of power with BigTech in a recent publication. He draws attention to a very unfavorable constellation that threatens our open liberal democracies: accumulation of capital, ownership of ubiquitous and central digital infrastructures and online platforms, the domination of public discourse and journalism, the control over persons and their data, and finally monopolies in AI science, technology and innovation. In order to protect and strengthen Western Liberal democracies in the Age of AI and the core trinitarian idea of 'human rights, rule of law and democracy', Nemitz argues, we need "a new culture of technology and business development …which we call human rights, rule of law and democracy by design". I will come back to this suggestion concerning design, since I think it is our only hope left.

Trust in the institutions of liberal-democracy is in decline all over the world[2], while independent media, freedom of expression and alternative sources of information are among the democratic parameters that have shown the greatest global decline in recent years. So what are the main threats to democracy? The first potential threat concerns the rise of political micro-targeting, which allows political actors to target individual citizens directly in their online browsing, so that every user gets a different, specifically tailored advertisement to see. This tailored advertising model is a key ingredient of the digital economy and a core features of the online platform environment. But it is potentially disastrous for democratic politics when political advertisements remain in the 'dark', or are unknowable to anyone except the sender, receiver, and the digital platform intermediary. At that point, customized political advertisements erode the public dimension of democratic politics. Different political actors should be able to contest one another's ideas in a public sphere, open to all.

The second problem is that the rise of a 'new online public space' can be harmful when the 'old' shared public space is left vacant, or when it leads to further 'filter bubble' divisions and balkanization. Citizens may focus on a private online world whereas a public sphere withers away. This gives a completely new meaning to Habermas' structural change of the public sphere.

A third harm can come from disinformation, which is intentionally misleading information, provided for political purposes and gain. A fourth related harm might come from 'algorithmic distortion', where the selection of content by algorithms online might enhance political filter bubble divisions and echo chambers.

[2] Annual Report of the Varieties of Democracy Institute (V-Dem, 2018); Democracy Index (The Economist Intelligence Unit, 2017); (Foa and Mounk, 2017).

In the age of ubiquitous digital technology our values, ideals and principles need to be designed for. For a long time it has been believed that new communication technologies could be developed in order to increase the quality and frequency of the public deliberation required for a healthy democracy. Yet, relatively few online spaces up till now have been able to realize this widely advertised potential of the Internet. There are hardly any good examples of online environments that encourage citizens to deliberate together about political issues of great importance in ways that help them to appreciate that their disagreements are based on different experiences and cultures, values and perspectives, and support them to bring their awareness to bear effectively on their deliberations and decision-making.

In fact, currently, the vast amount of debate that goes on in on-line discussions is characterized by the dominance of trending topics, volatile sentiments, filter bubbles, group-think, bias, cascades of false information, lack of information and cognitive distortions, break-ups and interruptions, both glaring and subtle errors of fact and reasoning. Are there perhaps features that make online communication poorly suited to developing meaningful civic or political relationships? Does the on-line context, when it is not carefully designed, encourage uncivil discourse, does it facilitate diffusion of unverified information, does it polarize opinions and does it facilitate the balkanization of groups. Klein observes that although extant large-scale online argumentation platforms do "provide unprecedented opportunities for interacting on a massive scale", they yet have "to realize their potential for helping people deliberate effectively, typically generating poorly-organized, unsystematic and highly redundant contributions of widely iying quality". (M. Klein)

Democratic structures and the new digital era

In the digital era, an important component of the elaboration of democratic structures is the ensemble of processes that lead to the making of standards, and the extent to which they can be promoters of the so-called "public interest". Interoperability plays an essential role in ensuring such structures and thus in promoting the public interest.

Indeed, without common or shared standards, interoperability is impossible. This became evident in the 1960s, when computers were unable to communicate with each other due to each manufacturer implementing its own proprietary protocols. The ARPANET project was born to overcome this problem, by creating an open network and ending the era of closed architectures that trapped users within one system. With the advent of the Internet as a "network of networks" in the 1980s, thanks to the TCP/IP protocols, diversity in computer networks was enabled. However, the trend towards closed systems remains, as demonstrated by the formation of "silos" in digital social platforms (Cromity and Stricker, 2011), or the limitations of interoperability in e-books (Sire, 2021).

Digital standards today reflect the history of opening and closing strategies, geopolitical and economic interests, and the technical approaches taken by various actors such as private sector leaders, librarians, researchers, and other stakeholders in network governance. Lawrence Lessig stated, "code is law": it is also true that standards are laws –particularly for consumers, who can be trapped in closed systems due to divergent standardization policies. These policies can determine the success or failure of a system, and affect its capability to serve as a truly consensual and/or democratic arena for the expression of public opinion. The battles over standards – such as those focused on computer networks in the 1990s (Russell, 2014) or on videotex (Flichy, 1998) and its famous achievement, the French Minitel (Mailland and Driscoll, 2017) – demonstrate that digital standards are not just technical issues, but rather relationships of power and negotiation. As such, they mirror evolving economic and geopolitical interests, including the public interest, while also being rooted in a longer term-perspective. Digital standards are a technical, social, economic, and political construct.

Among the interests mirrored by standards and standard-making processes are issues of defining and promoting the public interest, and by extension, democratic principles. Alison Harcourt and colleagues point out the importance of asking about standard-setting organizations questions such as how these organizations promote public interest, the extent to which corporate goals are balanced with respect to it, and whether decision-making procedures account for the ability of consumer and digital rights groups to defend citizen/consumer rights (Harcourt et al., 2020: 5).

Democratic principles affected by new technologies[9],[10]

As already described, new technologies have an impact on existing democratic principles. The advent of the digital age has brought about a profound transformation in democratic structures across the globe, offering a plethora of unprecedented opportunities while simultaneously posing considerable challenges. As digital technologies continue to evolve at a rapid pace, they are reshaping the ways in which citizens engage with political processes, disseminate information, and operate democratic institutions. These technologies have the potential to enhance democratic engagement in a number of ways. They can facilitate communication between citizens and their representatives, increasing access to political discourse. They can also foster inclusivity through social media and online platforms. The digital connectivity that is emerging promises to empower voices that have been traditionally marginalised, thereby providing new avenues for political participation and civic activism.

However, as previously indicated in the abstract, the advent of the digital era also introduces a series of profound risks that threaten the integrity of democratic processes and institutions. The proliferation of misinformation, the ascendance of algorithmically-driven echo chambers, and the commercialisation of online spaces present considerable challenges to the democratic ideals of pluralism, fairness, and informed decision-making. Furthermore, these technologies have the potential to exacerbate existing societal divides, reinforcing existing biases and limiting exposure to diverse perspectives. In light of these broader implications of digital technologies for democracy, a more detailed analysis of their specific impact on core democratic institutions is warranted.

The functioning of any democracy is contingent upon three fundamental pillars: an independent judiciary, a free press, and the protection of freedom of speech. As democracy grapples with the challenges of the digital age, these institutions are subjected to new pressures that could erode their foundational roles. For example, the judiciary may be susceptible to the influence of public opinion amplified through social media, which could potentially compromise its impartiality. The free press, which has long been regarded as the "fourth estate" in democratic systems, is facing growing challenges from the proliferation of fake news and disinformation, which are eroding journalistic standards and public trust. Concurrently, the right to freedom of speech, a fundamental tenet of democratic discourse, is a subject of contention in the context of the internet, where the necessity to regulate harmful content coexists with the imperative to safeguard the right to open expression. These institutions are not only affected by the general trends of digital disruption, but also face unique challenges that require targeted solutions.

Democratic Principles

Democracy is a complex system, comprising a set of core principles and standards that provide the framework for its operation. These democratic fundamentals not only inform the processes of political decision-making but also exert an influence on the structures of society. In recent years, technological advancements have significantly altered the landscape of democratic engagement, prompting a re-evaluation of traditional practices and the emergence of new forms of political participation.

The core principles of democracy entail the involvement of the citizenry in the political process. Indeed, active citizen participation constitutes a fundamental tenet of democracy. This encompasses both direct engagement in political processes and representation through elected officials. It is imperative that effective participation be ensured in order to guarantee the inclusion of a multiplicity of voices in the decision-making process. The principle of political equality necessitates that the votes of all citizens be accorded equal weight in the electoral process. This is of vital importance in guaranteeing that all individuals are afforded an equitable opportunity to influence the processes of governance. The rule of law requires that legislation be applied in a consistent manner to all individuals, including those in positions of authority within the government. The principle of transparecy serves to foster accountability and uphold justice within the legal framework, thereby ensuring the fair and just administration of justice. Transparency in government actions is of paramount importance in fostering public trust. The accessibility of information to the general public enables an examination of the decision-making processes of those in positions of authority, thereby facilitating the holding of those in power to account. Those elected to office are duty-bound to respond to the public for their actions. The continued trust in democratic institutions is contingent upon the efficacy of accountability mechanisms. The protection of fundamental human rights is an indispensable aspect of democracy. It is of the utmost importance that all individuals are able to engage in political processes without fear of discrimination, in order to ensure the continued health of democratic societies. It is of the utmost importance that elections are conducted impartially in order to guarantee genuine competition among candidates. It is of the utmost importance to maintain the integrity of electoral processes in order to preserve public confidence in democracy. A flourishing democracy is characterised by a respect for the rights of minorities and dissenting voices, which together foster an environment conducive to debate and allow for a diversity of opinions.[3]

An Independent Judiciary

The authors highlight the potential for digital technologies to undermine the independence of the judicial system. They also draw attention to the possibility that the influence of public opinion and the activities of online influencers could compromise the impartiality of the judiciary. However, the use of digital technologies in legal disputes also presents a

[3] Council of Europe, Twelve Principles of Good democratic governance, see https://www.coe.int/en/web/centre-of-expertise-for-multilevel-governance/12-principles; Council of Europe, Reykjavík Principles for Democracy, see https://www.coe.int/en/web/steering-committee-on-democracy/10-principles-for-democracy (accessed Nov. 24, 2024).

potential risk that must be considered. It is essential to ensure that citizens do not perceive a lack of human involvement in the judicial process.

A democratic judiciary system is based on an impartial, fair, and constitutionally guaranteed system of courts of law with independent and professional judges[4]. Impartiality of judicial rulings means they should be taken on the facts of each case, including factual evidence, individual merits and credibility, legal arguments, and relevant laws. Fairness is understood as the fair treatment of people in the administration of justice. The growing adoption of AI as a tool for supporting the administration of justice in the legal system is primarily driven by the efficiency, accuracy, and flexibility of AI in many effort-intensive tasks. At the same time, the improper use of AI, or even the outright substitution to humans, poses inherent risks to the fairness, transparency and explainability of the decisions supported by the AI models, as well as to the transparency and accountability of the decision-making process (CEPEJ, 2018). Next, we explore the applications, advantages, and risks of using AI in a democratic judiciary system.

Applications of AI to law have been envisaged since the late 1970's. Early projects, as surveyed by (Cook et al., 1981), addressed: information retrieval from legal documents; automatic drafting of legal documents; formalization and reasoning over rules in the legal domain; understanding the patterns of legal argumentation and of the decision-making procedures of attorneys. The pivotal special issue of the AI & Law journal (Sartor and Branting, 1998) focused more specifically on judicial applications of AI, distinguishing two practical goals of AI: building tools to support judicial activities, and developing new analytical tools for understanding and modeling the judicial process. In that special issue, (Schild, 1998, Tata, 1998, Leith, 1998) provided a critical review of software systems for sentencing support since the 1980's, broadly categorizing them as: statistical systems, which provide aggregated historical data; rule-based systems, which encode expert knowledge; and case-based systems, which look for similar cases to a given one. In the 1990's, the "statistical systems" were not yet AI-based[5], and the burden of comparing case/offender features with historical aggregates was left on the judges. This was a major weakness as judges do not have the required data analysis skills to conduct such a comparison. Other defects pointed out by (Schild, 1998) for statistical systems include: the inability to verify factual accuracy, the lack of statistical confidence, the incompleteness of the historical database, and the usage of biased/unrepresentative collection of historical data. (Tata, 1998) criticized the reductionist approach of computing similarity of cases based on the classification of doctrinal categories. (Leith, 1998) argued that the formal logic approach used in expert systems is in contrast with the discretion of judges, and advocated for AI designers to understand the role of judges and their needs. Interestingly, all these weaknesses are nowadays considered potential biases that AI systems can be subject to (Alvarez et al., 2024).

The rise of concerns about fairness in Machine Learning (ML), a branch of AI, gained widespread attention in 2016 following criticism by ProPublica (Larson et al., 2016) against the COMPAS algorithm used for recidivism prediction. The use of ML in the criminal justice system is covered in the survey by (Berk et al., 2021). The authors point out the lack of conceptual precision in the discussions of fairness in criminal justice risk assessments. They review six kinds of fairness notions, some of which are inherently incompatible with one another and with the objective of maximizing accuracy. For instance, the COMPAS algorithm fails to meet an equal false positive rate among social groups, but it achieves equal calibration (Corbett-Davies et al., 2017). That is the result of different perspectives and moral values taken by the designers of the algorithm and the ProPublica journalists. In essence, AI algorithms often adopt an utilitarian approach, prioritizing outcomes that benefit the majority over the minority. This sacrifices fairness at an individual level, as noted by (Forrest, 2021). A second major issue traces back to the roots of ML: the historical data used for training ML models to accomplish a certain task, e.g., to predict recidivism of defendants. The data used for training is assumed to encode the ground "truth" of the task, e.g., the actual outcome of recidivism for each defendant in case the defendant would have been released. In the analysis of the COMPAS algorithm, ground truth was approximated by the actual re-arrest outcome of defendants in the two years period after they were scored. However, due to the unobservability of crime, re-arrest does not coincide with re-offense (Bao et al., 2021), which is the recidivism outcome intended to be predicted. Furthermore, we do not know whether or not defendants who were not released would have recidivated in case they would have been released (Alvarez et al., 2024). These considerations show the difficulty of collecting ground "truth" data.

The recent technological breakthroughs in generative AI have boosted the applicability of AI in the judicial domain (Lai et al., 2023), as well as increased the acceptance and perceived efficacy of AI-powered robot lawyers (Xu et al., 2022). Large Language Models (LLMs), in particular, enhance the understanding and generation of human-like language with unprecedented accuracy and versatility, also in specialized settings such as the legal domain. These enhanced capabilities have led to advancements in several applications, including legal document summarization, contract analysis and generation, legal question answering, legal text classification, and legal reasoning (Anh et al., 2023). For example, with regard to contract reviews, the study by (Martin et al., 2024) found that LLMs show comparable performances as Junior

[4] See https://www.principlesofdemocracy.org (accessed Nov. 24, 2024).

[5] Interestingly, (Schild, 1998) includes them in the 'Non-Intelligent Systems' section of the paper.

Lawyers and Legal Process Outsourcers, but at much faster speed and lower cost, thus "challenging the status quo and calling for a reimagined future of legal workflows". However, it is worth noting that LLMs lack an explicit reasoning mechanism akin to logic-based approaches. Moreover, they may perpetuate biases present in the training data, potentially resulting in unfair outcomes. Lastly, LLMs could pose privacy concerns by potentially exposing sensitive information contained within the training data.

Trustworthy AI research is becoming of utmost importance for ensuring not only fairness but also human agency and oversight, accountability, explainability, robustness and safety, privacy, diversity, reproducibility, and societal and environmental well-being (Kaur et al., 2023). In particular, eXplainable AI (XAI) has been rapidly growing in the last decade, also in the context of judicial systems. While symbolic AI and statistical approaches from the 1980s were inherently interpretable, modern state-of-the-art AI models are too complex or large for humans to clearly understand the rationale behind specific outcomes. As a result, post-hoc explanation methods have become necessary, tailored to the specific needs of different users. For example, XAI approaches can assist judges in decision-making, aid litigants in persuading judges, and empower defendants to question AI-based decisions. (McGregor Richmond et al., 2024) highlight the distinction between explanations as justifications for exercising authority and discretion, and explanations as mechanisms for understanding the inner workings of a system. The former includes establishing a connection between the facts considered and the ultimate decision made. The latter is the subject of technical research in the XAI field, which remains relatively young and currently faces several limitations (Alvarez et al., 2024). For instance, multiple explanations may exist for a specific outcome of an AI model, potentially leading to disagreement among the parties regarding the reasons for the model's output. Furthermore, post-hoc explainability methods, often relying on a surrogate interpretable model of a black box, are not always guaranteed to be stable or faithful to the underlying black box, resulting in inaccurate explanations.

Free Press and freedom of speech

The paper identifies the deterioration of journalistic standards, driven by the spread of misinformation and the commercialisation of online spaces, as a significant threat to the integrity of a free press. As previously stated in the introduction, digital platforms also present a complex challenge to freedom of expression. It is necessary to achieve a delicate balancing act between the imperative of free speech and the growing risks posed by harmful content and disinformation.

Description and background

The contemporary context of content circulation online is suffering from a democratic paradox. Society lives in what might appear as a golden age of freedom of expression, as it has never been so easy to make an idea public, and to share it with the greatest numbers of people. However, at the same time, never in history has the power to constrain information – to limit, filter, block it – been so readily available via largely private concentration points in the digital public sphere.

Both scholarship and public policy have largely focused content moderation discussions on the role and power of the large human-facing platforms of social media and the Web. These platforms already shape the public sphere in platform design decisions and in how their algorithms order information flows. Far from the early Internet ideals of a "free flow of information," it is now a routine practice for their moderation systems, merging autonomous detection and human supervision in opaque ways, to sometimes empower and sometimes constrain and punish what can or cannot be said. Large platforms face a dueling set of economic incentives – an interest in enforcing norms of behavior online versus an underlying Faustian revenue model exchange largely free services in exchange for the monetization of personal user data.

How do new technologies influence the freedom of speech and free information?[14],[15]

Since their beginnings, the vast majority of online discussion spaces have used a number of moderation devices and systems, without them being accused of arbitrary censorship. Indeed, the largely shared definition of censorship implies a State-driven control of public expression, aiming at allowing or preventing the diffusion of specific discourses in public debate. Moderation, instead, is concerned with private and limited spaces, and the rules that prevail on a website are not necessarily applicable to others. Content control in these spaces is about applying a private regulation typically contained in a user agreement, whose primary aim is most often to pacify conversations and exchanges, and whose legitimacy is generally not questioned by these users. If platforms such as Facebook, Youtube or X (the formerly Twitter) are accused today of engaging in censorship practices, it is because they have become privileged sites for the democratic debate to take place, "infomediaries" that are crucial tools and places of daily discussion and mobilization. Thus, the limits between

an “internal” regulation and what pertains to the structure and shared norms of public debate have become much more uncertain and opaque in the last ten to fifteen years.

The power of these platforms also lie in delegation, the new forms of public-private partnerships that filter public opinion or deputize these firms to carry out content control on behalf of the state. Several laws addressing content control online in the late 2010s and early 2020s (such as the “Avia” law on freedom of expression in France, or the NetzDG law in Germany) have thus been criticized for basically the same reason, i.e. the fact that private companies act in place of judges in a heavily opaque context. Laws that require platforms to remove flagged content within 24 hours entail a risk of “over-censorship”, inasmuch as companies prefer to block legitimate content rather than risk incurring large fines. Moreover, what constitutes false information or an expression of hate speech is subject to interpretation; and the two categories are vague enough in several contexts to open up forms of political instrumentalization of online content control and moderation.

In this scenario, the core issue appears to be no longer – as it may have appeared in earlier days of the Internet – to know whether online content should be moderated or not, but who has to regulate and how. This debate today extends far beyond the relations between States and private actors; it now includes a variety of actors including journalists, activists, companies, developers and “ordinary” Internet users; yet, the multistakeholder procedures and organizations that have been developed to address regulation in other areas of Internet governance are so far largely absent.

Freedom of Speech and Privacy

The principles of freedom of speech and privacy are considered to be fundamental tenets of democratic societies. They are enshrined in international human rights frameworks as being essential for the autonomy of the individual and the health of public discourse. In the digital age, however, these principles face unprecedented challenges, as the very technologies that enable instantaneous global communication also facilitate new forms of surveillance, censorship, and manipulation. The advent of digital platforms, social media, and vast data networks has ushered in an era of hyper-connectivity, enabling individuals to disseminate their views on a far greater scale than was previously possible. Nevertheless, this enhanced connectivity has also obscured the demarcation between free speech and privacy, as both state actors and private enterprises exercise considerable influence over online domains.

On the one hand, the digital age has facilitated the democratisation of expression, enabling voices that have previously been marginalised to be heard and expanding the public sphere. However, as previously discussed, this increased openness has also facilitated the proliferation of harmful speech, including disinformation, hate speech, and extremism. This has led governments and digital platforms to implement content moderation strategies that may inadvertently suppress legitimate expression. In countries with authoritarian regimes, digital tools have been employed as instruments of control to suppress dissent. This has involved the use of surveillance technologies to monitor and silence critics. Even in countries with democratic governments, the question of how to balance the need to safeguard national security with the right to freedom of expression has become increasingly challenging. This is exemplified by the use of spyware such as Pegasus to monitor journalists and political opponents.

Concurrently, the diminution of privacy in the digital domain represents a considerable menace to the freedom of expression. It is common practice for online platforms to collect vast amounts of personal data, frequently without users being fully aware of the extent of the data being collected or having given their consent to this. This data is then used to create detailed profiles that can be exploited for commercial, political, or governmental purposes. The practice of surveillance, conducted by both state and non-state actors, has the effect of undermining the conditions necessary for the exercise of free speech. This is because individuals may be dissuaded from expressing their views online, due to a fear that their activities are being monitored. This "chilling effect" on speech is further compounded by the dearth of transparency and accountability in the manner of personal data collection, dissemination, and utilisation.

Even though it might seem that freedom of speech pertains to the public sphere and the right to privacy pertains to the private sphere, it turns out that both rights are two sides of the same coin (Nyst, 2018). In our days, there is automatic monitoring and censoring of anti-government websites (e.g., in China), tapping of the phones of journalists and political opponents (Morocco, Bahrain, and a host of countries including even some Western countries like Spain, via the Pegasus software), surrendering of user data to the law enforcement agencies (Google), suppression of posts regarded as subversive by the local authorities (Facebook), etc.

The above actions affect the right to freedom of opinion and expression (Article 19 of the Universal Declaration of Human Rights) and the right to private life and private communications (Article 12 of the Declaration). In fact, each of those two rights is a prerequisite for the other, as pointed out by the UN Special Rapporteur on freedom of opinion and expression (LaRue, 2013). If you know you are being observed by some hostile or untrusted entity, you do not speak freely; rather,

you will most likely self-censor. Hence, privacy versus undesired intrusion is necessary for free speech to be real. Conversely, if you do not enjoy free speech, you are forced to disguise your real feelings and opinions even when talking to those in your private circle; hence your private life is degraded.

Thus, privacy-preserving legislation and technologies are necessary to make free speech happen.

Emerging risks in the digital age from the state and limitation of fundamental rights?

In the context of the digital age, it is of paramount importance to acknowledge the emerging risks associated with state actions that curtail fundamental rights through the utilisation of digital products. The equilibrium between harnessing technology for the public benefit and safeguarding individual liberties is a tenuous one, necessitating the establishment of robust legal frameworks and ethical guidelines. It is of the utmost importance to guarantee that states are held to account for how they utilise technology. This is a fundamental step in ensuring the continued safeguarding of democracy and the maintenance of public trust in governance.

As discussed in the introduction, the increasing use of social media, and reliance on it as a source of news and information on current affairs, has the potential to fragment and destabilise political discourse. Shared values and narratives are challenged and undermined, calling into question our models of democratic government and the legitimacy of public institutions. A possible effect, as apparent from the rise of populism in Europe and elsewhere, is that these developments generate a counterreaction from supporters of the existing order, who in their eagerness to re-assert firmer control over public discourse, resort to undemocratic means of political repression to do so. Here, the direction of risks emanating from digital technology shifts – it is no longer a vehicle for non-state actors to undermine democratic order, represented by the state, but rather the state itself that uses it as a means to foreclose political criticism and government accountability.

Today, the ability of digital technology to gather data and track the activities of citizens in real time, both on-line (by monitoring electronic communications and/or placing cookies and forms of spyware on computers) and off-line, through ubiquitous cameras and other sensors, has never been greater. In addition, recent advances in data analytics, including sophisticated AI-assisted algorithms, allow the data to be mined for information ever more revealing of an individual's interests and preferences, including predictions of their likely future conduct. It is apparent such developments pose a very serious challenge to democracy – for while the defensive reliance on such technologies may allow state institutions to survive attacks from outside, the internal values that made the state democratic in the first place – in particular tolerance of social and political diversity and respect for the private sphere and right to dissent of its citizens – will be lost in the process.[6] There is also the temptation for the state to use surveillance techniques 'positively' in the interests of more efficient government, including by tracking and rewarding behaviour on the part of citizens seen as exemplary.[7] There, too, the effect is to constrict liberal values, producing a cautious and dull consensus on how to speak and behave and what goals to pursue, so as to avoid losing out on benefits and advantages provided by the state.

The link between state surveillance and totalitarianism is historically well-documented.[8] It was the fear of the possibilities offered by computers to collect and store data on citizens, that in the 1970s prompted the first data protection codes and legislation, both at international and state level. An early example is the 1981 Data Protection Convention passed by the Council of Europe and which is binding on its signatory states.[9] In Article 1, data protection law was recognized as a necessary amplification of the traditional right of privacy in the era of computerised data processing: "The purpose of this Convention is to secure in the territory of each Party for every individual, whatever his nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him".

A further milestone was the 1983 decision of the German Constitutional Court in the so-called 'Census Case',[10] in which it held that informational self-determination (the right to choose how data concerning oneself is used) was an aspect of the fundamental right to personality under the German Basic Law. Otherwise, developments in the digital sphere, making it possible to gather, store and analyse unlimited amounts of personal data, carried the risk of seriously inhibiting the free

[6] On the danger of democratic collapse from within, see (Levitsky and Ziblatt, 2018).

[7] The Chinese State is generally perceived in the West as pursuing such a 'social scoring' model; see (Creemers, 2018).

[8] Notoriously, every authoritarian regime, from Metternich's Austria to Soviet Russia and beyond, has deployed state informants and secret police to monitor the activities and opinions of its citizens.

[9] Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Treaty No 108, Strasbourg, signed on 28 January 1981.

[10] Volkszählungsurteil des Bundesverfassungsgericht, 15 December 1983, BVerfGE, vol. 65, 14.

development of the individual personality: "A social order in which individuals can no longer ascertain who knows what about them and when and a legal order that makes this possible would not be compatible with the right to informational self-determination. A person who is uncertain as to whether unusual behaviour is being taken note of at all times and the information permanently stored, used or transferred to others will attempt to avoid standing out through such behaviour."[11]

This decision proved influential in many other European countries, and today data protection is regulated at EU level by the General Data Protection Regulation, which entered force in 2018.[12] The latter imposes various requirements upon data controllers – be they state or private actors - who accumulate data of natural persons, including the need for data to be obtained by lawful and fair means (where appropriate, with the individual data subject's consent), and for it to be relevant, accurate, and not excessive in relation to the purposes of use; furthermore, the data subject is equipped with various rights, including to find out what data is being held and used in relation to him, and to object, in defined circumstances to its further retention and use.[13]

Nevertheless, it remains the case that the risks to democracy presented by digital technology, including mass surveillance, are harder to guard against, when the state is the perpetrator than where the transgressor is a private actor. This is because the state has the monopoly on making and enforcing the relevant laws. So, while it may legally regulate and sanction the potentially undemocratic digital activities of non-state actors, as regards its own activities, much depends on the ethical attitudes and self-restraint of state actors themselves: are they prepared to forgo, for the sake of the higher public good, the advantages that digital surveillance could offer in enabling more efficient government, as well as defending the state from possible attack?

Here, rather than relying on the good conscience of individual political actors alone, a crucial role has been played by Liberal Constitutionalism, anchoring fundamental values and rights in 'higher' constitutional law (less easy to change through the legislative process than ordinary statutes), and appointing the judicial arm of government to serve as its guardians against the more volatile elected ('democratic' in the populist sense) arms of the legislature and executive. Thus, democratic states are constructed, according to the 'separation of powers' doctrine to avoid concentrating power in the hands of popular government, which might make repressive laws at the behest of a 'tyrannical majority'.[14] The European Union itself is modelled on these principles, with the Court of Justice ready to strike down secondary legislation enacted by the law-making institutions where it finds this to conflict with fundamental EU citizen rights enshrined in the EU Charter on Fundamental Rights. A good example, in relation to digital surveillance, is the Court's 2014 decision[15] to annul the EU Data Retention Directive (providing for the blanket collection and storage of individual telecommunication data for investigating potential terrorist offences) as a disproportionate and unjustified interference in citizens' rights to privacy and data protection under Articles 7 and 8 of the Charter. In a number of subsequent decisions, the Court has used similar reasoning in striking down data retention laws at member state level.[16]

In post-second world war Western Europe, this Constitutional Law approach to limiting state power has proven quite robust; however, it reaches its limits at the point at which a populist government becomes so powerful, and the underlying political climate so emotive and intolerant, that the role of the judiciary itself as a restraint on the legislature and executive is no longer accepted. In such a case, judges may be exposed to threats and intimidation, or the Constitution itself may be altered to attenuate the rights and freedoms it previously embodied. This underlines the need for democratic values to be a living part of democracy, rather than seen simply as potentially inconvenient words in a statute – this arguably requires the active and informed engagement of interested citizens in political discourse.

What are the risks and how can they be countered?

The risks can be both at the collective and individual level. Indeed, some technologies, such as electronic voting systems, might directly affect our society. Other perils to be considered are related to a more personal sphere, for example cyberbullying.

[11] Excerpt taken from the translation of the case by the Konrad-Adenauer-Stiftung, available at: https://freiheitsfoo.de/census-act/ (accessed Nov. 24, 2024).
[12] Regulation (EU) 2016/679.
[13] Ibid., Chapters II and III.
[14] (Mill, 1859/1977); the 'separations of powers' doctrine is set out in Montesquieu's 'The Spirit of the Law' from 1748 (Montesquieu, 1748/1989).
[15] Joined Cases C-293/12 and C-594/12 Digital Rights Ireland and Seitlinger and Others.
[16] Most recently, in Joined Cases C-793/19 and C-794/19, SpaceNet and Telekom Deutschland.

Finally, some risks, such as polarisation, and presence of biases, d/misinformation, might start individually and then dangerously grow and affect also the general public. These risks are the majority and must be tackled before they find fertile ground, with specific countermeasures. Hence, there is the need to act promptly on both causes (e.g., d/misinformation) and effects (e.g., opinion/ideology polarization/radicalization, presence of echo chambers) of online polluted information environments, since they also reflect in the physical world.

Thus, in our Research Infrastructure (RI) there are methods for recognizing and measuring early signals of the polarization process on specific topics, allowing the development of targeted mitigation strategies (e.g., carefully designed information campaigns) or data-driven content moderation strategies. We will provide some examples of real applications in the SoBigData RI in Section 5.2.

We obviously take particular care of and we strongly rely on the European legal framework. For example, we are confident in the application of the new instruments offered by the Digital Service Act, that obligates Very Large Online Platforms and Very Large Online Search Engines, including X, Meta, and Google, to establish mechanisms "to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content" (Article 16(1)), and to grant independent researchers that meet the specific requirements listed in Article 40(8), the so-called "vetted researchers", to request for data access for the specific researches.

Technical Considerations to address the threats to democratic principles

While democracy in the digital era comes with the many issues described in the previous sections, we have the possibility to develop technical countermeasures to recognize (possible at a very early stage) potential problems and to mitigate such problems, for example highlighting social bots that spam in online social networks, limiting echo chambers, discovering fake news, in order to offer a practical help in contrasting the threats, for example fighting d/misinformation with tailored campaigns.

In the following we analyse the technological point of view on such threats. In particular, in Section 5.2, we aim at offering an overview of solutions that are already developed within the SoBigData consortium to help governors and policy makers in dealing with these problems (van Biezen et al., 2012).

Technologies that can benefit democracy

Internet voting denotes electronic voting (e-voting) (Lipset and Rokkan, 1967) systems that allow votes to be cast using the internet. There are, however, other types of e-voting, like those based on optical ballots, those using computers without remote connection or those sent by phone (Kersting and Baldershein, 2004; Tula, 2005). All these systems can be used in public or private elections.

When using internet voting in a democratic framework, the same requirements placed on traditional paper ballot voting must be satisfied: the suffrage must be at least, universal, free, equal, and anonymous (Mitrou et al., 2002).[17]

There has been a socio-political debate on the pros and cons of internet voting for democracy. Among the cons, the digital divide stands out as an important issue (Norris, 2001): not only who can have internet access, but also who is familiar with the use of digital technologies. There is also another argument about technophobia: some people mistrust the internet not only for voting, but for anything else (like using credit cards or doing online shopping). Other people associate casting their vote with social interaction at the polling station, and they do not want to lose that.

Yet, there are many reasons to argue that e-voting can reinforce democracy in our digital age. Comfort for voters is one of them, but not the only one. More important is the *increased participation* enabled by information technology both in terms of quantity and quality (Braun, 2005). Internet voting offers more chances for citizens to participate in elections, especially for those living abroad, in isolated areas, or with health conditions that prevent them from reaching a polling station. Regarding quality of participation, it is also increased thanks to the amount of information available on the internet on the candidates and their programs (certainly, there is a danger of misinformation or bias in the available information, but this can also affect traditional paper ballot elections). Finally, e-voting reduces the cost of election, it is more sustainable from the environmental point of view, and it facilitates more frequent voting processes, which may enable deploying direct democracy (referenda) in many issues.

[17] See (Barrat et al., 2007) for a detailed discussion of those requirements.

The Role of SoBigData Research Infrastructure [24]

Within the SoBigData RI, we study the problems related to data democracy under different aspects, offering research solutions to deal with these threats. In particular, we have a whole research space dedicated to Societal debates and misinformation, and we offer more than one hundred resources in our Catalogue. In the following, we present an overview of the topic discussed, from bot detection (Mazza et al., 2022) to echo chamber detection (Morini et al., 2021), from analysing diffusion processes and opinion dynamics (Rossetti et al., 2018; Pansanella et al., 2023) to misinformation studies and fake news discovery (Zubiaga et al., 2018; Mayank et al., 2022).

Mazza et al. (2022), studied the difference, in terms of profile characteristics, activities (e.g., retweeting patterns) and shared content (e.g., massive use of URLs sharing), from humans, trolls, and social bots, i.e., automated accounts designed to impersonate humans. This is particularly important because the latter two categories of users are often involved in campaigns to influence public opinion, as well as to spread fake news, and conspiracy theories. Moreover, even the EU AI Act draft highlights the need to be completely transparent in the use of bots, meaning that it must be clear to users that they are interacting with bots instead of humans.

Rossetti et al. (2018) provided a library with models and methods for analysing diffusion processes, such as epidemics and opinion dynamics. It also offers an environment for the execution and the simulation of experiments and some online visualisation tools that abstracts its programmatic interface and makes available the simulation platform to non-technicians. Morini et al. (2021) studied the echo chamber phenomenon, in which beliefs are amplified or reinforced by communication repetition inside a closed system, i.e., subset of users in a network who share the same ideology and tend to have dense connections primarily within the same group. Here, the novelty of the proposed approach is to rely on an in the middle approach between the micro-scale of analysing single user's digital traces and the macro-scale of looking only at polarised communities.

The paper of Pansanella et al. (2023) analysed the polarisation phenomenon related with media biases, i.e., biases reported in media coverage, and reported the results of a model that simulates the shifting of the public's opinion concerning a specific topic and the increasing of extreme opinions when readers are exposed to mass propaganda. Interestingly, one of the authors' findings is that the power to push individuals towards the media opinion is not dependent on such opinion but on the open-mindedness of the population, interpreted as the aptitude of people to change their minds, thus, the more "open-minded" is the population, the easier agents conform around the promoted opinion(s).

Zubiaga et al. (2018) proposed a survey on rumours diffusion in online social networks, in particular on the classification between newly emerging rumours, i.e., not observed before, which need to be automatically detected using natural language processing (NLP) and data mining techniques, and long-standing rumours that are usually known but that circulate for long periods of time, which are sometimes very hard to debunk and for which the analysis is often based on a non real-time setting. Mayank et al. (2022) combines NLP and Knowledge Graph entities used to map named entities present in news titles and, thus, to discover fake news with great accuracy even with minimal amount of information. This enables the early detection of fake news even when it is difficult to extract or to have access to comments and other metadata of articles.

All this work will help the researcher to discover and target possibly early-stage problems, and treat them before they become a real issue for our democracy.

Legal Considerations to address these threats

At root, the risks posed by digital technologies to democratic values and processes stem from the ease and speed with which they enable the generation and dissemination of contentious information. Accepting this requires a departure from the view of philosophers and political theorists since The Enlightenment. They began with a strong presumption that the free flow of information was a good thing for society – it was seen as providing the essential foundation for considered political debate, by allowing a careful balancing of the relevant interests affected by a given decision. In this context, some thinkers regarded even the truth or falsity of the information as of secondary importance. As John Stuart Mill famously declared, the suppression of both kinds injures the public: "If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error."[18]

[18] (Mill, 1859, p. 19), as noted by (John, 2019), Mill's overall view was in fact more complex.

This attitude was reflected in the law, which in the context of political debate was reluctant to penalise the dissemination of information, including that which might prove false: in many countries members of parliament enjoyed immunity from liability for defamation in relation to their utterances in the legislative chamber, and even in the private sphere, courts would rarely stop publication in advance; rather, those injured by false speech had to rely on bringing proceedings after the event (Kirtley, 2020). More generally, free speech was also seen as a crucial element in individual self-development, and something the state should allow its citizens as a matter of natural law and the social contract (John, 2019). In the modern era, freedom of expression has featured prominently in all the great charters of human rights.[19] Later, in the optimistic aftermath of Western liberal democracy's victories against totalitarian regimes, some political theorists attributed these successes – in the spirit of Mill - to the openness of liberal societies to dissent and criticism: this enabled them to identify mistakes and eradicate inefficiencies more rapidly than 'closed' totalitarian systems (Popper, 1945).

The reason for setting out the above, is to call attention to how dramatically matters have changed as a result of the arrival of modern information technology. The old assumption that information per se is a good thing for democracy no longer seems to hold true once a certain saturation point has been reached. In particular, when there is too much information, and/or it is too complex or speculative, persons have difficulty in processing it (to decide how far to accept its validity and/or assess its implications for them). The danger is that, as a means of self-preservation, they turn to simpler, populist accounts of 'the truth'; accounts which are more straightforwardly linear in form, and with an - easily processed - emotive appeal. Such accounts are typically undemocratic in spirit, often seeking to scapegoat particular individuals or social groups in a dangerous and socially divisive manner.

In attempting to combat these dangers from contentious information in the digital age, the European Union lawmaker has singled out internet information intermediaries, especially operators of large social media fora, as the most efficient regulatory target. A recent example is the 2022 Digital Services Act,[20] which imposes stronger content monitoring duties upon such actors, especially the largest platforms (VLOPs)[21]. The hope thereby is that harmful information, including that tending to incite social division or otherwise subvert the democratic order, will be more swiftly identified and removed, thereby limiting its damaging effect. As Recital 3 of the Act states, "Responsible and diligent behaviour by providers of intermediary services is essential for a safe, predictable and trusted online environment and for allowing Union citizens and other persons to exercise their fundamental rights guaranteed in the Charter of Fundamental Rights of the European Union…., in particular the freedom of expression and information and…the right to non-discrimination." Platforms failing to supervise their content responsibly face penalties in the form of substantial fines.

It is undoubtedly true that the algorithms used by social media platforms, which in pursuit of advertising revenue, prioritise simpler and emotive opinions and news items (that their viewers are more likely to click on) over more sophisticated and nuanced accounts, have exacerbated the problem of antidemocratic speech attaining prominence. Arguably, though, as previously hinted, this may be more of a symptom than a cause, reflecting a deeper underlying problem: the oversaturation of public discourse with complex and fragmentary information that has caused many persons to lose their bearings. Here, applying restraints to social media outlets is unlikely to solve the problem (the appetite of a significant proportion of the public for simpler narratives, particularly those serving to reinforce their identity vis a vis scapegoated outsider-groups); the audience will seek and find such narratives elsewhere.

Accordingly, proactive initiatives to promote and encourage healthy democratic discourse are also imperative. This is indeed recognized by the EU in its 2023 Declaration on Digital Rights and Principles for the Digital Decade.[22] Chapter II, entitled 'Solidarity and Inclusion' affirms that: "Technology should be used to unite, and not divide, people. The digital transformation should contribute to a fair and inclusive society…", and commits the EU to "making sure that the design, development, deployment and use of technological solutions respect fundamental rights, enable their exercise and promote solidarity and inclusion". The question remains, though, as to how to give effect to this aspiration in practice. Arguably, besides technical solutions, further concrete approaches are needed to encourage citizens to re-engage in more traditional forms of political debate, as mediated by mainstream, democratic political parties. In most Western democracies, the major political parties have suffered significant declines in their membership over recent decades (van Biezen et al., 2012). In the process, an important regulatory function they perform is being lost, namely their tendency to temper division by drawing their support - if they are to have electoral success - from diverse sections and groups in society. As noted by political theorists, this cross-cutting function helps bridge social cleavages, and encourages mature political reflection and debate among their members, requiring them to find common ground with fellow members of divergent background (Lipset and Rokkan, 1967). In this regard, the EU, in cooperation with member state parliaments, might usefully consider

[19] Most famously in the First Amendment of the US Constitution (1791); in Europe it is protected by Article 10 of the ECHR (1953) and Article 11 of the EU Charter (2000).
[20] Regulation (EU) 2022/2065.
[21] 'Very Large Online Platforms' - defined as having at least 45 million active monthly users in the EU.
[22] European Declaration on Digital Rights and Principles (2023/C 23/01).

ways to incentivize citizens to become members of mainstream parties, rather than relieving their political frustrations in acrimonious social media debate.

Conclusion

The advent of the digital age has brought with it a duality of opportunities and challenges with regard to the protection of freedom of speech and privacy. While technological advancements have expanded platforms for expression, enabling a more inclusive and diverse public discourse, they have also paved the way for increased surveillance, censorship, and the erosion of personal privacy. As this chapter has demonstrated, the very digital tools that facilitate the empowerment of voices, particularly those from marginalised communities, can also be exploited by governments and corporations for the purposes of monitoring, silencing and manipulating speech. The emergence of content moderation practices, frequently opaque and inconsistently applied, further complicates the delicate equilibrium between the protection of free expression and the curbing of harmful content.

Furthermore, the extensive gathering and utilisation of personal data by governmental and non-governmental entities present considerable threats to individual autonomy and freedom of expression. The fear of being surveilled, tracked, or targeted has the effect of deterring individuals from engaging in public discourse, either through self-censorship or a complete withdrawal from the discourse. This represents a significant challenge to the very essence of democratic participation, which is contingent upon the freedom to express ideas without fear of reprisal or intrusion.

In order to ensure the preservation of democratic values in the digital age, it is of the utmost importance that policies and regulations are devised and implemented with the objective of safeguarding both the right to freedom of expression and the right to privacy. This necessitates a collaborative approach involving governments, technology companies, civil society, and international bodies to establish transparent and accountable systems that safeguard individual rights. It is of the utmost importance to guarantee that content moderation is conducted in a fair, transparent, and respectful manner that upholds the right to free speech, while simultaneously limiting the dissemination of harmful content. It is equally important to implement more robust privacy protections, including restrictions on data collection and surveillance practices, to guarantee that individuals can participate in online discourse without the constant fear of monitoring.

In conclusion, the advent of the digital age has redefined the boundaries of freedom of speech and privacy, compelling societies to re-examine the manner in which these fundamental rights can be safeguarded in an increasingly interconnected world. By confronting these challenges directly and implementing well-considered, rights-based policies, it is possible to ensure that the advantages of digital technologies are exploited while safeguarding the freedoms that are indispensable to the well-being of democratic societies.

References

Alvarez JM, Colmenarejo AB, Elobaid A et al. (2024) Policy advice and best practices on bias and fairness in AI. Ethics Inf Technol, 26, 31. https://doi.org/10.1007/s10676-024-09746-w

Anh DH, Do D-T, Tran V, Minh NL (2023) The Impact of Large Language Modeling on Natural Language Processing in Legal Texts: A Comprehensive Survey. KSE, (pp. 1–7). https://doi.org/10.1109/KSE59128.2023.10299488

Bao M, Zhou A, Zottola S et al. (2021) It's COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks. NeurIPS Datasets and Benchmarks. https://doi.org/10.48550/arXiv.2106.05498

Barrat I Esteve J, Castellà-Roca J, Domingo-Ferrer J, Reniu I Matamala JM (2007) Internet voting. In Encyclopaedia of Digital Government, (pp. 1125–1159). Hershey PA: Idea Group. https://doi.org/10.4018/978-1-59140-789-8.ch170

Barrett B, Dommett K, Kreiss D (2021) The capricious relationship between technology and democracy: Analyzing public policy discussions in the UK and US. Policy Internet, 13, (pp. 522–543). https://doi.org/10.1002/poi3.266

Berk R, Heidari H, Jabbari S, Kearns M, Roth A (2021) Fairness in criminal justice risk assessments: The state of the art. Sociological Methods & Research, 50(1), (pp. 3–44). https://doi.org/10.1177/0049124118782533

van Biezen I, Mair P, Poguntke T (2012) Going, going,…gone? The decline of party membership in contemporary Europe. European Journal of Political Research, 51, (pp. 24–56). https://doi.org/10.1111/j.1475-6765.2011.01995.x

Braun N (2005) E-voting–worldwide developements, opportunities, risks and challenges. In Reflections on the future of democracy in Europe (pp. 115–119). Strasbourg: Council of Europe.

Bundesverfassungsgericht (1983) Census judgement of the Federal Constitutional Court (Volkszählungsurteil des Bundesverfassungsgericht), 15 December 1983, BVerfGE, vol. 65, 14.

CDDG–European Committee on Democracy and Governance, Mergel I (2021) Study on the Impact of Digital Transformation on Democracy and Good Governance. 64 pages. Strasbourg: Council of Europe.

CEPEJ–European Commission for the Efficiency of Justice (2018) European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. 77 pages. Strasbourg: Council of Europe.

Congge U, Guillamón M-D, Nurmandi A, Salahudin, Sihidi IT (2023) Digital democracy: A systematic literature review. Front. Polit. Sci., 5:972802. https://doi.org/10.3389/fpos.2023.972802

Cook S, Hafner CD, McCarty LT, Meldman JA, Peterson M, Sprowl JA, Sridharan NS, Waterman DA (1981) The applications of artificial intelligence to law: a survey of six current projects. AFIPS National Computer Conference 1981, (pp. 689–696). https://doi.org/10.1145/1500412.1500516

Corbett-Davie S, Pierson E, Feller A et al. (2017) Algorithmic decision making and the cost of fairness. KDD, ACM, (pp. 797–806). https://doi.org/10.1145/3097983.3098095

Council of Europe (1953) European Convention on Human Rights, Article 10. Rome: Council of Europe.

Council of Europe (1981), Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Treaty No 108, Strasbourg, signed on 28 January 1981.

Council of Europe, Reykjavík Principles for Democracy, https://www.coe.int/en/web/steering-committee-on-democracy/10-principles-for-democracy (accessed Nov. 24, 2024).

Council of Europe, Twelve Principles of Good democratic governance, https://www.coe.int/en/web/centre-of-expertise-for-multilevel-governance/12-principles (accessed Nov. 24, 2024).

Court of Justice of the European Union (2014) Judgment (Grand Chamber). Joined Cases C-293/12 and C-594/12, 8 April 2014.

Court of Justice of the European Union (2022) Judgment (Grand Chamber). Joined Cases C-793/19 and C-794/19, 20 September 2022.

Creemers R (2018) China's Social Credit System: An Evolving Practice of Control. 32 pages. http://doi.org/10.2139/ssrn.3175792

Cromity J, De Stricker U (2011) Silo persistence: It's not the technology, it's the culture!. New Review of Information Networking, 16(2), (pp. 167–184). https://doi.org/10.1080/13614576.2011.619924

The Economist Intelligence Unit (2017) Democracy Index 2017 – Free speech under attack. 78 pages. London/New York/Hong Kong: The Economist Intelligence Unit.

European Parliament and Council (2022) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, L 277, (pp. 1–102).

European Union (2000) Charter of Fundamental Rights of the European Union, Article 11. Official Journal of the European Communities, C 364, (pp. 1–22).

European Union (2023) European Declaration on Digital Rights and Principles. Official Journal of the European Union, C 23, (pp. 1–10).

Flichy P (1998) La normalisation : un processus d'explication du travail technique. Le cas des caractères du vidéotex. Réseaux, Communication – Technologie – Société, 16(87) (pp. 105–116). https://doi.org/10.3406/reso.1998.3164

Foa RS, Mounk Y (2017) The Signs of Deconsolidation. Journal of Democracy, 28(1), (pp. 5–15). https://doi.org/10.1353/jod.2017.0000

Forrest KB (2021) When Machines Can Be Judge, Jury, and Executioner: Justice the Age of Artificial Intelligence. 160 pages. Singapore: World Scientific. https://doi.org/10.1142/12172

Harcourt A, Christou G, Simpson S (2020) Global Standard Setting in Internet Governance. 290 pages. Oxford: Oxford University Press.

John RR (2019) Freedom of expression in the digital age: a historian's perspective. Church, Communication and Culture, 4(1), (pp. 25–38). https://doi.org/10.1080/23753234.2019.1565918

Kaur D, Uslu S, Rittichier KJ, Durresi A (2023) Trustworthy Artificial Intelligence: A Review. ACM Comput. Surv., 55(2), 39, (pp. 1–38). https://doi.org/10.1145/3491209

Kersting N, Baldershein H (2004) Electronic voting and democracy: A comparative analysis. 309 pages. London: Palgrave Macmillan. https://doi.org/10.1057/9780230523531

Kirtley JE (2020) Uncommon Law: The Past, Present and Future of Libel Law in a Time of "Fake News" and "Enemies of the American People". University of Chicago Legal Forum, 2020, Article 6.

Lai J, Gan W, Wu J, Qi Z, Yu PS (2023) Large Language Models in Law: A Survey. CoRR abs/2312.03718. https://doi.org/10.48550/arXiv.2312.03718

Larson J, Mattu S, Kirchner L, Angwin J (2016) How We Analyzed the COMPAS Recidivism Algorithm. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed Nov. 24, 2024).

LaRue F (2013) Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. United Nations, A/HRC/23/40.

Leith P (1998) The Judge and the Computer: How Best Decision Support? Artif. Intell. Law, 6, (pp. 289–309). https://doi.org/10.1023/A:1008226325874

Levitsky S, Ziblatt D (2018) How Democracies Die. 320 pages. New York: Crown.

Lipset SM, Rokkan S (1967) Party systems and voter alignments: Cross-national perspectives. 553 pages. New York: Free Press.

Mailland J, Driscoll K (2017) Minitel – Welcome to the Internet. 240 pages. Cambridge, MA: The MIT Press.

Martin L, Whitehouse N, Yiu S, Catterson L, Perera R (2024) Better Call GPT, Comparing Large Language Models Against Lawyers. CoRR abs/2401.16212. https://doi.org/10.48550/arXiv.2401.16212

Mayank M, Sharma S, Sharma R (2022) DEAP-FAKED: Knowledge graph based approach for fake news detection. Accepted at IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2022. https://doi.org/10.48550/arXiv.2107.10648

Mazza M, Avvenuti M, Cresci S, Tesconi M (2022) Investigating the difference between trolls, social bots, and humans on Twitter, Computer Communications, 196, (pp. 23–36). https://doi.org/10.1016/j.comcom.2022.09.022

McGregor Richmond K et al. (2024) Explainable AI and Law: An Evidential Survey. Digital Society, 3, Article 1. https://doi.org/10.1007/s44206-023-00081-z

Mill JS (1859) On Liberty. 207 pages. London: John W. Parker and Son.

Mill JS (1977) On Liberty. In Robson JM (Ed.), The Collected Works of John Stuart Mill (Vol. 18). Toronto: University of Toronto Press.

Mitrou L, Gritzalis D, Donos P, Georgaroudi G (2002). Legal and regulatory issues on e-voting and data protection in Europe. Financed from Eu-IST-2000-29518. Mytilene: University of the Aegean.

Montesquieu C-L de Secondat (1748) De l'esprit des lois. 868 pages. Geneva: Barrillot & Fils.

Montesquieu C-L de Secondat (1989) The Spirit of the Laws. Transl. a. ed. by Cohler AM, Miller BC, Stone HS. 759 pages. Cambridge: Cambridge University Press.

Morini V, Pollacci L, Rossetti G (2021) Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study. Appl. Sci. 11:5390. https://doi.org/10.3390/app11125390

Norris P (2001) Digital divide: Civic engagement, information poverty and Internet worldwide. 303 pages. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164887

Nyst C (2018) Two sides of the same coin – The right to privacy and freedom of expression. Privacy International. https://privacyinternational.org/blog/1111/two-sides-same-coin-right-privacy-and-freedom-expression (accessed Nov. 24, 2024).

Pansanella V, Sîrbu A, Kertesz J, Rossetti G (2023) Mass media impact on opinion evolution in biased digital environments: a bounded confidence model. Sci Rep, 13:14600. https://doi.org/10.1038/s41598-023-39725-y

Popper KR (1945) The open society and its enemies. 721 pages. London: Routledge.

Rossetti G, Milli L, Rinzivillo S, Sîrbu A, Pedreschi D, Giannotti D (2018) NDlib: a python library to model and analyze diffusion processes over complex networks, International Journal of Data Science and Analytics, 5(1), (pp. 61–79), Springer International Publishing. https://doi.org/10.1007/s41060-017-0086-6

Russell AL (2014) Open Standards and the Digital Age: History, Ideology and Networks. 306 pages. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139856553

Sartor G, Branting K (1998) Introduction: Judicial Applications of Artificial Intelligence. Artif. Intell. Law, 6, (pp. 105-110). https://doi.org/10.1007/978-94-015-9010-5_1

Schild UJ (1998) Criminal Sentencing and Intelligent Decision Support. Artif. Intell. Law, 6, (pp. 151–202). https://doi.org/10.1007/978-94-015-9010-5_3

Sire G (2021) Le dernier refuge, Essai sur la standardisation du livre numérique. Mémoire d'habilitation à diriger des recherches. HDR, Paris: Université Panthéon-Assas.

Tata C (1998) The Application of Judicial Intelligence and Rules to Systems Supporting Discretionary Judicial Decision-Making. Artif. Intell. Law, 6, (pp. 203–230). https://doi.org/10.1007/978-94-015-9010-5_4

Tula MI (2005) Voto electrónico. Entre votos y máquinas. las nuevas tecnologías en los procesos electorales. 396 pages. Barcelona: Ariel.

United States (1791) First Amendment of the US Constitution.

V-Dem–Varieties of Democracy (2018) Democracy for All? – V-Dem Annual Democracy Report 2018. 95 pages. Gothenburg: University of Gothenburg, Department of Political Science.

Xu N, Wang K-J, Lin C-Y (2022) Technology Acceptance Model for Lawyer Robots with AI: A Quantitative Survey. Int. J. of Social Robotics, 14, (pp. 1043–1055). https://doi.org/10.1007/s12369-021-00850-1

Xu Z (2023) Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities, Applied Artificial Intelligence, 36(1):2013652. https://doi.org/10.1080/08839514.2021.2013652

Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R (2018) Detection and Resolution of Rumours in Social Media: A Survey. ACM Comput. Surv. 51(2) Article 32 (February 2018), 36 pages. https://doi.org/10.1145/3161603