

# Report on the First Ethical Web Science Workshop (EWSW 2025) at The ACM Web Science Conference 2025 (WebSci 2025)

Oshani Seneviratne

Rensselaer Polytechnic Institute  
Troy, NY, USA  
senevo@rpi.edu

Marc Spaniol

Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR6072  
Caen, France  
marc.spaniol@unicaen.fr

John S. Erickson

Rensselaer Polytechnic Institute  
Troy, NY, USA  
erickj4@rpi.edu

Gaël Dias

Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR6072  
Caen, France  
gael.dias@unicaen.fr

## Abstract

The Web Science conference series (WebSci) is an established venue for Web related research in an interdisciplinary field dedicated to understanding the complex and multiple impacts of the Web on society and vice versa. This is where the Ethical Web Science Workshop kicks in—by bringing together quantitative, qualitative and mixed methods research, including techniques from the social sciences and computer science. The Workshop aims at investigating ethics in research and Web-based data collections. To this end, explainability is a core aspect of Ethical Web Science. Systems failing to deliver data without providing accessible and meaningful to stakeholders risk perpetuating opacity and eroding trust. Hence, explainability is not only a technical problem but a fundamental ethical issue. Thus, the workshop investigates the shift in global digital discourse from a technical and societal point of view.

**Date:** 20 May, 2025.

**Website:** <https://ethical-web-science.github.io>.

---

# 1 Introduction

The Web Science Conference 2025 (WebSci25) “Maintaining a Human-Centric Web in the Era of Generative AI” was a venue for cross-disciplinary research where the Web itself is the primary object of research. In 2025, it was held at Rutgers University, New Brunswick, New Jersey, USA. The Ethical Web Science Workshop was the first of its kind and was seeking to explore the foundational ethical challenges posed by the rapidly evolving relationship between the Web and cutting-edge technological advancements, particularly in AI. Having reached a point where ethics of artificial intelligence receives a lot of attention, it is equally important to obtain a deeper understanding of the ethical foundations of the Web which served as an origin of this development. While Generative AI, large language models (LLMs), recommender systems, and other transformative solutions are often powered by massive data sets sourced from the Web, they are commonly generated from user-generated contents that have never been intended for that purpose. For that purpose, the Ethical Web Science workshop has been launched in order to investigate principles of fairness, transparency, and consent in the context of Web data exploitation. Moreover, the workshop provided a timely opportunity to connect Web Science ethics to ongoing debates surrounding data governance, computational social science reproducibility, and the responsibilities of researchers leveraging increasingly autonomous AI-driven Web ecosystems.

Further information about the program of the Ethical Web Science Workshop is available at the Workshop Website<sup>1</sup>.

## 2 Objectives

The main objective of the Ethical Web Science Workshop at WebSci’25 aims to address the ethical challenges that arise from the rapidly evolving relationship between the Web and AI. With increasing reliance on Web-sourced data, issues of fairness, transparency, and consent have become central to technological innovation. We address issues such as data provenance, explainability, and digital commons in the context of ethical research and technology.

To this end, topics of the Ethical Web Science Workshop were addressing ethical challenges in Web Science and included, but were not limited to the following:

- Explainability and Accountability
- Data Provenance and Consent
- Algorithmic Amplification and Harm
- Web Fragmentation and Inclusivity
- Responsibility in Web Research
- Redefining Digital Commons

In addition to these core topics, the workshop explicitly encouraged submissions that combined empirical studies with normative analysis, recognizing that ethical Web Science must integrate methodological rigor with value-sensitive design.

---

<sup>1</sup><https://ethical-web-science.github.io>

---

### 3 Contributions

For its first edition, the Ethical Web Science Workshop was organized as a half-day workshop. Three very positively reviewed submissions were accepted for oral presentation [Seneviratne et al., 2025b] covering various aspects at the intersection of the Web, AI, and ethics. In addition, we were proud to receive two distinguished keynotes by Dominic DiFranzo (Lehigh University, Bethlehem, PA, USA) and Fred Morstatter (University of Southern California, Los Angeles, CA, USA).

The workshop drew participants from academia, industry, and the public sector, reflecting the wide-ranging relevance of ethical questions surrounding Web data. Discussions throughout the workshop highlighted not only current challenges but also emerging opportunities for collaboration across disciplines and institutions.

The Ethical Web Science Workshop was introduced by Oshani Seneviratne on behalf of the Workshop organizers by giving a quick intro of the workshop objectives and the program of the day [Seneviratne et al., 2025a]. The session then started with the first keynote by Dominic DiFranzo on “We Must Save Us: Rebuilding Social Media Research for a Safer Digital Future”. In his talk he investigated the shifting landscape of social media research in the year 2025. To this end, he addressed how community-driven approaches might help to overcome the problems arising political pressure and public skepticism. As such, he pointed out that research in that direction will be more important than ever in order to sustain this research field under constrained conditions. The first paper presentation was then given by M. A. Mridul, I. Kang, and O. Seneviratne (Rensselaer Polytechnic Institute, Troy, NY, USA) on “Terminators: Terms of Service Parsing and Auditing Agents” [Mridul et al., 2025]. This paper dealt with problems arising from the often lengthy and complexly written Terms of Service (ToS) documents. They introduced a modular framework called Terminators employing LLMs, which aim at improving the understandability of ToS documents. To this end, they developed an approach based on a comprehensible three-step process of term extraction, verification, and accountability planning in order to ensure greater transparency instead of a black-box summarization. By doing so, they aim at empowering users to better understand their digital rights. The session was concluded by the paper on “Bursting the Filter Bubble with Knowledge Graph Inversion” contributed by F. Spadea and O. Seneviratne (Rensselaer Polytechnic Institute, Troy, NY, USA) [Spadea and Seneviratne, 2025]. In their paper, the authors addressed the issues arising from ethical concerns when recommender systems lead to a reinforcing filter bubble. For that purpose, they propose to incorporate a personalized knowledge graph (KG) that preserves privacy and disrupts filter bubbles at the same time. To this end, they employ knowledge graph edge inversion in order to provide alternative viewpoints and stimulate diversity.

The second session then started with the keynote by Fred Morstatter on “Understanding and Mitigating Bias in Networked Language Games and Social Platforms.” To this end, he explored the emergence and amplification of bias in networked digital interactions through a sequence of experiments and methods. Starting with a controlled network experiment on shaping consensus he then extended this framework by introducing LLMs highlighting how their outputs reflect and influence bias within social interactions. Continuing with with studies of real-world social media platforms, he concluded his presentation by introducing a retrieval-augmented generation (RAG) framework aiming at identifying and contextualizing these threats. Subsequently, the paper of R. Zhao and J. Wright (University of Oxford, Oxford, UK) on “Introducing an Auditing Layer to Web Science” was presented [Zhao and Wright, 2025]. In their talk the authors introduced a conceptual

---

design of a framework of data-harvesting Web Science practices, employing a common auditing layer. In particular, they addressed the framework’s practical, scientific, and ethical advantages in an era of LLMs, autonomous agents, and wide-scale prevalence of AI in general. As such, they considered it as a viable technical solution, fostering discussions on ethical while low-burden data collection and processing across the Web. The workshop was then concluded by closing remarks of James Hendler (Rensselaer Polytechnic Institute, Troy, NY, USA). who emphasized the importance of the discussions carried out in this inaugural workshop.

## 4 Conclusion and Future Directions

The inaugural Ethical Web Science Workshop at WebSci’25 demonstrated both the urgency and the growing momentum behind examining the ethical foundations of Web-based research and technologies in an era increasingly shaped by generative AI. The workshop attracted a diverse group of scholars across computer science, social science, and policy, fostering lively discussion on topics ranging from explainability and data provenance to algorithmic amplification, auditing infrastructures, and the responsibilities of researchers working with large-scale Web data.

Several important lessons emerged. First, ethical challenges on the Web cannot be addressed solely through technical means; sustained interdisciplinary engagement is essential to uncover the social, legal, and methodological assumptions embedded in Web data practices. Second, explainability and transparency remain central pillars for trustworthy Web technologies, especially as generative models become tightly intertwined with Web-scale content. Finally, the workshop underscored strong community interest in building practical frameworks, such as auditing layers, and user-empowering tools for understanding digital rights, that advance ethical Web Science.

Given the enthusiastic participation and the breadth of topics addressed, there is clear interest in continuing this conversation in future years. Participants expressed strong interest in developing a shared set of conceptual and technical guidelines for ethical Web data collection and analysis. Several attendees proposed follow-up collaborations on topics such as shared auditing infrastructures, ethics-aware data pipelines, and comparative studies of consent practices across Web platforms. These emerging directions provide a foundation for planning subsequent editions of the workshop. Thus, subsequent editions of the Ethical Web Science Workshop may explore additional themes including responsible Web data curation for AI training, governance models for autonomous Web agents, standardization of ethics-aware auditing mechanisms, and cross-cultural perspectives on Web norms and consent. By cultivating an ongoing forum for critical reflection and methodological innovation, the workshop aims to contribute to a more transparent, accountable, and human-centric Web ecosystem.

## Acknowledgements

The Ethical Web Science Workshop 2025 thanks the authors of submitted papers for their efforts, and all workshop attendees. The Ethical Web Science Workshop was jointly organized by Rensselaer Polytechnic Institute (USA) and University of Caen Normandy (France). We also wish to express our particular thanks to our program committee members:

- 
- Tanya Singh (Rensselaer Polytechnic Institute, USA)
  - K. Krasnow Waterman (LawTechIntersect LLC, USA)
  - Gautam Kishore Shahi (University of Duisburg-Essen, Germany)
  - Kathy Fontaine (Rensselaer Polytechnic Institute, USA)
  - Dirk Ahlers (Norwegian University of Science and Technology, Norway)

Last but not least, we would like to thank the organizers of WebSci'25 for hosting the workshop.

## References

- Maruf Ahmed Mridul, Inwon Kang, and Oshani Seneviratne. Terminators: Terms of service parsing and auditing agents. In *Companion Publication of the 17th ACM Web Science Conference 2025*, WebSci Companion '25, page 44–48, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715358. doi: 10.1145/3720554.3736183. URL <https://doi.org/10.1145/3720554.3736183>.
- Oshani Seneviratne, Marc Spaniol, John S. Erickson, and Gaël Dias. Ethical web science workshop overview. WebSci Companion '25, page 38, New York, NY, USA, 2025a. Association for Computing Machinery. ISBN 9798400715358. doi: 10.1145/3720554.3737366. URL <https://doi.org/10.1145/3720554.3737366>.
- Oshani Seneviratne, Marc Spaniol, John S. Erickson, and Gaël Dias, editors. *1<sup>st</sup> Ethical Web Science Workshop, New Brunswick, NJ, USA, May 20, 2025*, 2025b. ACM. ISBN 9798400715358.
- Fernando Spadea and Oshani Seneviratne. Bursting the filter bubble with knowledge graph inversion. In *Companion Publication of the 17th ACM Web Science Conference 2025*, WebSci Companion '25, page 39–43, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715358. doi: 10.1145/3720554.3736182. URL <https://doi.org/10.1145/3720554.3736182>.
- Rui Zhao and Jesse Wright. Introduce an auditing layer to web science. In *Companion Publication of the 17th ACM Web Science Conference 2025*, WebSci Companion '25, page 49–53, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715358. doi: 10.1145/3720554.3736184. URL <https://doi.org/10.1145/3720554.3736184>.