

Privacy notice for research related on polarization, inauthentic campaigns, and mental well-being on Bluesky

This privacy notice pertains to the collection and processing of data from Bluesky for scientific research purposes. Bluesky (<https://bsky.app/>) is a publicly accessible decentralized social media platform that allows users to publicly share short posts, images, and engage in public conversations.

We are committed to transparency in our data collection practices and to protecting the privacy rights of Bluesky users. This notice outlines how we collect, use, store, and protect the data obtained from Bluesky, in compliance with applicable data protection laws and regulations.

1. What is being researched in this study and what is the purpose of processing personal data?

This study aims to explore and analyze polarization patterns and coordinated information campaigns on the Bluesky social media platform. For polarization, we quantify the levels of division across various topics using advanced network science-based methods and other statistical techniques. By examining these patterns, we believe to gain valuable insights into how opinions and discussions evolve and diverge within the Bluesky community, and use these findings as a proxy for drawing wider societal conclusions. We will also investigate what is the role of mental well-being in polarization by analysing the content of the public messages. Additionally, inauthentic online campaigns can manipulate people's opinions, amplify extreme views or even threaten social cohesion. By constructing networks based on public interactions and posts on Bluesky and analyzing them, we can estimate the extent of inauthentic campaigns and coordinated behavior in today's digital spaces.

This research is conducted in the context of three projects: DECA*, CON-NET**, and POLEMIC***.

**) The Democratic Epistemic Capacities in the Age of Algorithms (DECA) project examines how information is used and impacted across societal levels, from individuals to institutions and digital infrastructures.*

****) The Integrating Content and Network Structure for Detecting, Understanding, and Mitigating Online Misbehavior (CON-NET) project focuses on identifying trends, signals, and suspicious behaviors as well as misbehaving entities shaping the flow of information and polarization patterns in online domains.*

*****) POLEMIC: POLarization Entangled with Mental well-being: Integrated and Computational approach to quantifying the underlying feedback loop. The project focus on understanding how political polarization and mental well-being are related.*

2. What personal data is processed in the study?

By personal data, we mean your Bluesky *account name* and *public profile*, as well as your *posts*. If you participate in any public discussion on Bluesky, you may be an indirect participant in this research project. We collect public posts and public account information from participants in these discussions.

A *post* includes content publicly visible on the Bluesky platform (i.e., the content of the post, the time of publication, and the username of the poster or reposter). We do not have access to your private information (e.g., private posts or direct messages).

Processing of special categories

The personal data we collect may fall into special categories depending on the content you post. For example, if your posts are related to political topics, we are likely to collect information about your political opinions. For example, your views on government policies related to climate politics would fall into a special category of data (i.e., political opinions).

Sources of Personal Data

We collect the personal data described above from the public API (<https://docs.bsky.app/>) provided by Bluesky.

3. Necessity of personal data processing and removal of identifiable information

In this study, we only process information that is necessary for the purpose and implementation of the research. During the analysis phase of the study, the research data is primarily processed in a form where your identity is not directly identifiable to researchers without additional information.

Your Bluesky username will be replaced with a random identifier in our internal data processing and analysis. An exception to this is made when necessary: The initial data collection phase is the exception, where usernames are retained to ensure we don't duplicate data. Further, in some of the coordination and bot detection algorithms use user names input.

When sharing research data with research collaborators, your username will always be pseudonymized with a random identifier. It will not be possible to identify you from the scientific publications or other published research results derived from this study.

4. Legal basis for processing personal data

- The basis for processing personal data is scientific research in the public interest
- The basis for processing personal data is the explicit consent of the research subject

5. Sharing of personal data

Research data containing your personal information in pseudonymized format may be disclosed to third parties, such as scientific publications, for the peer review process of scientific publishing or for other purposes essential for the validation or publication of scientific research results.

The collected research data may be used for further scientific research in the same field or for research in other scientific fields supporting the study. As part of potential follow-up research, the research data may also be disclosed to other universities or research institutions if necessary for conducting the research.

Again, the research data is provided to collaborators in a form where direct identifiers such as names (in this case, Bluesky usernames) have been removed or replaced with random identifiers. This means it is not possible to directly identify a research subject from the research data without additional information. We can also share just the

internal IDs used by the Bluesky for profiles and posts, which means that the collaborators will need to refetch the data from Bluesky and any removed posts and profiles will not be visible to them.

6. International Data Transfers

Data containing personal information from the research project will not be transferred outside the European Union (EU) or the European Economic Area (EEA), or to international organizations.

7. Storage and Protection of Personal Data

Information processed in information systems

Your personal data is processed and stored in secure systems approved by Aalto University. The computers and information systems used are protected by username and strong password. Access to research material containing personal data is technically restricted so that only individuals participating in the research and those essential for conducting the research have access to your personal data.

Where and for how long the material is archived:

Fully anonymized data may be archived, for example, in the Zenodo service (<https://zenodo.org/>).

8. Storage and Deletion of Personal Data

The collected research data may be used for further research in the same scientific field or for research in other scientific fields supporting the study. Research material containing personal data will be deleted five (5) years after the last scientific publication that has utilized the research material in question.

9. Rights of Research Participants

According to the General Data Protection Regulation (GDPR), individuals have the right to:

- 1) Receive information about the processing of their personal data
- 2) Access the data concerning them
- 3) Request correction of their data
- 4) Request restriction of processing of their data
- 5) Object to the processing of their data
- 6) Have their data erased, if the conditions of Article 17(1) of the GDPR are met and the processing is no longer necessary for archiving purposes in the public interest, scientific research purposes or statistical purposes in accordance with Article 89(1)...

In the event that conducting the research does not require or no longer requires the identification of the data subject, the controller is not obliged to acquire additional information in order to identify the data or the data subject solely for the purpose of enabling the data subject to exercise their rights. If the controller cannot identify the data that can be linked to a particular data subject, the data subject does not have the right to inspect the data, correct the data, object to the processing of the data, or have the data erased. However, if the data subject provides additional information that allows them to be identified from the research data, these rights will not be restricted.

10. Controller's contact information

The controller of this research study is Aalto University Foundation sr., operating as Aalto University.

Person in charge of the research study

Questions regarding the conduct of the research study may be addressed to the person in charge of the study: Mikko Kivelä (mikko.kivela@aalto.fi)

Data Protection Officer

If the research participant has questions or requests related to data protection or the processing of personal data, the research participant should contact the Data Protection Officer of Aalto University: tel. +358 9 47001 (exchange), dpo@aalto.fi .

In this data request service, you can request the exercise of your rights under GDPR from Aalto University as the controller <https://datarequest.aalto.fi/en-US/>.

If a participant of the research study feels that his or her personal data has been processed in violation of data protection legislation, the participant has the right to lodge a complaint with the supervisory authority, the Data Protection Ombudsman's office (read more: <http://www.tietosuoja.fi>).