

# Sentiment Analysis with Large Language Models on Bluesky: Tag Groupings and Decentralized Social Media

Olga Grigorieva, Małgorzata Kurcjusz-Gzowska, Elen Muradyan, Suren Mnatsakanyan

Warsaw University of Technology

olga.grigorieva.stud@pw.edu.pl, malgorzata.kurcjusz.stud@pw.edu.pl,  
elen.muradyan.stud@pw.edu.pl, suren.mnatsakanyan.stud@pw.edu.pl

**Supervisor: Anna Wróblewska**

anna.wroblewska1@pw.edu.pl

## Abstract

Large language models can be used for sentiment analysis on social media, but most research focuses on centralized platforms such as Twitter and Facebook. Bluesky is built on the decentralized AT Protocol, where moderation and feed generation are modular and users' tagging practices are still evolving. These properties can affect both sentiment signals and how they interact with hashtag-driven discovery.

In this project we implemented an end-to-end experimental pipeline on Bluesky data, combining exploratory analysis of multiple feeds with targeted experiments on POLITISKY24. Concretely, we (i) profiled feed samples and POLITISKY24 with respect to volume, language tags, engagement, reply structure, and temporal activity; (ii) extracted hashtags and built tag-centric summaries and co-occurrence graphs; (iii) applied transformer-based sentiment and emotion inference (including confidence scores) and trained lightweight TF-IDF baselines (Logistic Regression, Multinomial Naive Bayes, LinearSVC) for comparison under weak supervision. These steps directly support our research questions on transfer to Bluesky (RQ1) and tag grouping (RQ2). Bias and fairness analysis (RQ3), calibration against gold labels, and robustness stress tests are defined as the next stage.

**Keywords:** large language models, sentiment analysis, decentralized social media, Bluesky, AT Protocol, hashtag grouping, hashtag clustering, multimodal sentiment, annotation practices, bias and fairness.

In this paper, we use the terms 'tag' and

'hashtag' interchangeably to refer to hash-tags used within social media posts.

## 1 Introduction

Large language models (LLMs) have changed sentiment analysis on social media. Earlier methods relied on lexicons or supervised classifiers for short, noisy texts, while transformers enable context-sensitive representations and domain-specific fine-tuning. GPT-style, LLaMA-family, and other open-source LLMs now deliver strong zero- and few-shot performance across various domains, including news, finance, health, and multilingual social media (Zhang et al., 2024). However, success depends on prompt design, domain alignment, and calibration. At the same time, the architecture of social media itself is shifting. Bluesky and the AT Protocol separate identity, hosting and feed generation, which gives users to move between providers and enables the operation of multiple custom feed generators and labeling services (Kleppmann, 2024). This decentralized design raises new questions: how do sentiment signals behave when feeds, labeling, and moderation are modular, and how does decentralization influence their flow and interpretation? The role of hashtags is also crucial, because they shape discovery, format topics, and help to identify the community. Previous work on centralized platforms has used deep learning and graph-based methods for tag recommendation, dynamic adaptation, and clustering (Djenouri et al., 2019; Liou et al., 2020). And recent studies leverage LLMs to refine topics and explain clusters, but they rather ignore decentralized platforms. Our goal in this project is to integrate these strands by implementing and evaluating LLM-based sentiment analysis methods specifically for Bluesky:

- explore existing Bluesky-native sentiment and tag datasets, including the Bluesky Social Dataset and POLITISKY24 (Rostami et

al., 2025; Failla et al., 2025), which include user-generated posts, political stance labels, and, where available, multimodal content.

- benchmark LLMs and standard transformer baselines on decentralized social media data;
- develop LLM-supported tag-grouping methods that fit the AT Protocol architecture;
- assess bias, fairness, and uncertainty when sentiment and tags are used in simulated Bluesky ranking pipelines.

The outcome will be a professional, reproducible framework for studying sentiment on decentralized social networks, along with concrete tools and datasets that other researchers can reuse.

## 2 Literature Review

### 2.1 Sentiment Analysis on Social Media

LLMs perform well on standard polarity classification, although dedicated architectures continue to perform better on structured tasks such as aspect-based sentiment analysis and opinion-role extraction (Zhang et al., 2024). Existing benchmarks show us that general-purpose LLMs can compete with fine-tuned transformers, especially when only small labeled datasets are available. Domain-specific work reflects this mixed picture. GPT-style and encoder-decoder models can match or outperform fine-tuned transformers with well-crafted prompts, but their performance drops on noisy or highly specialized material (He et al., 2024). Industry reports highlight the benefits of rapid, multilingual deployment, while also acknowledging ongoing challenges related to prompt design, safety, and operational costs. On platforms more similar to Bluesky, fine-tuned BERT, BERTweet and open LLMs boost political sentiment detection. Recent open models close much of the remaining gap when given enough in-domain data. Predictions are sensitive to linguistic issues such as emojis, sarcasm, code-switching, and non-standard varieties. Paraphrasing noisy posts can raise accuracy. However, it may also erase minority language forms. Multilingual studies show encouraging results with well-written prompts, although performance remains uneven for low-resource languages (Nasution, 2023; Fu, 2023) and this is important for Bluesky, which includes large English

and Japanese communities and is becoming more linguistically diverse (Sahneh et al., 2025).

### 2.2 Decentralized Social Media and Bluesky

Bluesky is built on the AT Protocol. It separates the social graph, identity, and content hosting, allowing providers to interoperate (Kleppmann, 2024). Moderation and feed curation are modular, allowing labeling services and feed generators to run independently. Early analyses of Bluesky’s growth point to fast uptake, varied posting patterns, relatively low toxicity, and active moderation, although these studies rely on classical toxicity metrics rather than LLM-based sentiment analysis (Sahneh et al., 2025). Broader research on decentralized protocols shows that decentralization redistributes, but does not eliminate, control over moderation or the structural inequalities tied to it (Huang, 2024).

### 2.3 Hashtags and Hashtag Groupings

Hashtags help to organise content, support discovery, and influence how topics and forming of communities. Deep learning models outperform bag-of-words methods for predicting hashtags (Djenouri et al., 2019), while approaches such as H-ADAPTS and dynamic graph transformers capture shifting usage patterns and infer new tags (Liou et al., 2020). Co-occurrence graphs and community-detection techniques reveal clusters linked to themes or actors, and LLMs can refine topic labels or reduce noise using clustering tools like BERTopic. Most existing work assumes centralized platforms with stable architectures, leaving hashtag grouping in decentralized, instance-specific environments largely unexamined (Feng et al., 2015).

### 2.4 Bias, Fairness and Multimodal Sentiment

Bias and fairness are central to LLM sentiment analysis. Fine-tuned models can produce systematic differences across demographic attributes despite achieving high accuracy (Radaideh et al., 2025), whereas adversarial training and post-hoc debiasing reduce bias on Twitter benchmarks (Venugopal et al., 2023). Decentralized networks’ normative features (blocklists, opt-in search) may amplify such biases in moderation or ranking (Huang, 2024). Multimodal sentiment adds complexity: images improve classification for short, neutral, or sarcastic text, but current models struggle with cultural references or text-image contra-

dictions and are often poorly calibrated (Jin et al., 2024). Annotation and uncertainty estimation are critical, as disagreements often reflect ambiguity, and access to images can shift labels (Kadriu et al., 2022). Calibrated models help flag unreliable predictions (Xiao et al., 2023).

### 3 Research Objectives and Questions

This project aims to develop and evaluate a comprehensive framework for LLM-based sentiment analysis on Bluesky, accounting for tag groupings, multimodality, where applicable, and decentralization.

#### 3.1 Objectives

- O1. **Dataset exploration:** Explore existing Bluesky post datasets, with the option to construct new datasets if existing ones are insufficient, ensuring high-quality human and LLM-assisted annotations that capture disagreement and uncertainty.
- O2. **Model evaluation:** Benchmark LLMs and transformer baselines in zero-shot, few-shot, and fine-tuned settings, including multilingual performance.
- O3. **Tag grouping:** Develop LLM-enhanced clustering and graph-based tag grouping methods that account for cross-instance and cross-feed variation.
- O4. **Bias and fairness:** Measure and mitigate demographic and political biases in LLM sentiment predictions.

#### 3.2 Research Questions

- **RQ1:** How well do LLMs generalize from centralized datasets to Bluesky in terms of accuracy, calibration, and robustness to platform-specific language and tags?
- **RQ2:** How can tag groupings be modeled with LLM embeddings and graphs, and how stable are they across instances and feeds?
- **RQ3:** What social or demographic biases appear in LLM sentiment predictions, and how effectively can mitigation techniques reduce them?

## 4 Proposed Methodology

Our methodology is organized around a pipeline that begins with dataset profiling and tag extraction, then proceeds to sentiment/emotion modeling and tag grouping analyses, and finally (in the

remaining work) evaluates calibration, robustness, and bias.

### 4.1 Data Collection and Preprocessing

In the current stage, we relied on existing Bluesky datasets and our collected feed samples. Specifically, we used the Bluesky Social Dataset and POLITISKY24 where applicable, and we treated feeds as separate samples to compare how content and metadata differ across generators. Text preprocessing includes basic normalization, language filtering when needed, and extraction of hashtags from post text/context. We also standardize timestamps to enable per-day activity summaries.

If additional coverage is required, we plan to extend the collection using the Bluesky firehose, with stratified sampling across time periods (e.g., major events), topics, and observable feed generators, while respecting user privacy and access constraints (e.g., private accounts are excluded).

### 4.2 Dataset Construction and Annotation

At this stage we primarily analyze existing labels (POLITISKY24 stance metadata) and create derived labels from model inference for exploratory purposes. In particular, we generate sentiment and emotion predictions together with confidence scores, and use these predictions to study tag-label relationships and tag groupings.

In the next stage, we plan to introduce a small human-annotated evaluation subset (or reuse any available gold labels) to obtain reliable estimates of accuracy and calibration on Bluesky-specific content. If multimodal coverage becomes feasible, we will extend the setup to text-image posts, following platform-specific annotation guidelines and recording annotator confidence to support uncertainty-aware evaluation.

### 4.3 Modeling

**Sentiment and emotion inference.** We apply off-the-shelf transformer models trained outside Bluesky to study transfer to Bluesky text. In our current implementation we used distilbert-base-uncased-finetuned-sst-2-english for sentiment inference and j-hartmann/emotion-english-distilroberta-base for emotion inference, storing both predicted labels and confidence scores.

**Baselines under weak supervision.** To compare against transformer inference with lower computational cost, we train classical TF-IDF baselines (Logistic Regression, Multinomial Naive Bayes, LinearSVC). These models are trained on a pseudo-labeled dataset produced by the sentiment transformer, using stratified splits and cross-validation for model selection.

**Tag groupings.** We extract hashtags and construct a tag co-occurrence graph. We then analyze tag usage conditioned on predicted sentiment/emotion labels and explore grouping tags using distributional similarity (e.g., clustering tag-label profiles) and graph structure. Stability across samples/feeds is treated as a key evaluation dimension for RQ2.

**Multimodal modeling.** Multimodal sentiment is part of the original project scope, but in the current stage we focus on text-only pipelines. Multimodal experiments (image-text fusion and uncertainty-aware calibration) remain planned future work.

#### 4.4 Bias, Fairness and Uncertainty

Bias and fairness analysis is scheduled for the next stage once we finalize an evaluation protocol with gold labels or controlled counterfactual probes. We plan to probe bias using paired inputs that vary demographic or identity markers and report group-level differences, including calibration gaps. Mitigation methods under consideration include reweighting, representation debiasing, and post-hoc calibration.

For uncertainty and calibration, the current notebooks store model confidence scores; the next step is to evaluate calibration against gold labels using metrics such as Expected Calibration Error (ECE) and reliability curves, and to introduce robustness tests focusing on Bluesky-specific phenomena (hashtags, slang, emojis, and short/no-context posts).

### 5 Results

This section summarizes the empirical outputs produced so far. Since part of the pipeline currently relies on model-inferred labels (pseudo-labels), we report results in two layers: (i) descriptive results from EDA and tag analysis, and (ii) modeling results from transformer inference and TF-IDF baselines. Where appropriate, we refer-

ence the corresponding notebooks to ensure traceability and reproducibility.

#### 5.1 Dataset profiling and EDA outputs

Across the collected Bluesky feed samples, we computed per-feed statistics covering: number of posts, number of unique users, language tag distributions (`langs`), word-count proxies derived from `text`, engagement signals (likes, reposts, replies), the fraction of posts with zero engagement, reply prevalence (`reply_to`), and daily activity time series. These results establish that feeds differ not only in volume but also in interaction patterns and language composition, which motivates treating “platform transfer” as more than a single test set evaluation (RQ1).

For POLITISKY24, we produced stance distributions conditioned on target entity, compared stance and confidence-level frequencies, analyzed text-length differences by stance/target, and extracted hashtags to compare the most frequent tags used for different political entities. These outputs provide a concrete view of Bluesky-specific political tagging practices and how they vary across subsets (RQ1, RQ2).

#### 5.2 Tag analysis and grouping signals

To connect labeling with tag usage, we applied an off-the-shelf emotion classifier (`j-hartmann/emotion-english-distilroberta-base`) and stored both predicted labels and confidence scores. Using extracted hashtags, we constructed a tag co-occurrence graph and summarized tag frequencies conditioned on predicted emotion labels. We additionally explored clustering based on tag-label profiles to obtain higher-level tag groupings. These results constitute the first implemented version of our RQ2 methodology: tag groupings induced by (i) co-occurrence structure and (ii) label-conditioned distributional similarity.

A key limitation is that current label-conditioned tag summaries rely on model predictions rather than human gold labels. Therefore, the current results are best interpreted as exploratory structure discovery that guides the design of the next-stage evaluation protocol.

#### 5.3 Sentiment modeling results (transformer inference + TF-IDF baselines)

We applied `distilbert-base-uncased-finetuned-sst-2-english` to Bluesky text

and stored predicted sentiment labels and confidence scores. We then trained TF-IDF baselines (Logistic Regression, Multinomial Naive Bayes, LinearSVC) on the pseudo-labeled dataset to evaluate how well inexpensive classical models can reproduce transformer decisions under weak supervision.

We report: (i) qualitative agreement patterns between baselines and transformer outputs, (ii) cross-validated performance summaries for classical models on the pseudo-label task, and (iii) distributional properties of model confidence scores as a precursor to calibration analysis. A full generalization and calibration evaluation against gold labels is planned as the next stage (RQ1).

## 6 Work Plan & Current Status

We keep the original plan, but report progress and the remaining steps explicitly. Our milestones are structured to answer RQ1-RQ3 with reproducible artifacts and clear evaluation protocols.

### 6.1 Completed milestones

- Established an analysis pipeline for Bluesky feed samples, including timestamp normalization and core EDA summaries (volume, users, language tags, engagement, replies, and activity over time).
- Performed EDA for POLITISKY24, including stance distributions by target entity, confidence-level analysis, text length comparisons, and hashtag extraction with entity-specific top-tag summaries.
- Implemented transformer-based inference for emotion labeling and produced tag graphs and tag-label summaries to support grouping analyses.
- Implemented sentiment inference and trained classical TF-IDF baselines (LogReg, MultinomialNB, LinearSVC), including stratified train/test splits and cross-validation for model selection under weak supervision.

### 6.2 Next milestones (to answer RQ1-RQ3 fully)

- Introduce a small human-annotated evaluation subset (or reuse any available gold labels) to measure true generalization on Bluesky (accuracy, macro-F1) and calibration (e.g., ECE / reliability curves).
- Add robustness checks focused on Bluesky-specific phenomena: hashtags, code-

switching, slang, emojis, and short/no-context posts.

- Formalize tag grouping stability tests across feeds/instances (graph community detection stability, embedding-based clustering stability).
- Execute bias and fairness analysis (RQ3) with explicit group definitions or counterfactual probes, then evaluate mitigation (reweighting, calibration, prompt constraints, or debiasing baselines).
- Integrate sentiment/tag signals into a simple feed simulation to study downstream social effects (optional extension).

### 6.3 Expected contributions

Key deliverables include (i) cleaned and documented Bluesky-derived datasets (or dataset subsets) with clear preprocessing and labeling procedures, (ii) benchmark results and reproducible code for transformer inference and TF-IDF baselines, (iii) methods and analyses for hashtag grouping and its stability across feeds/instances, and (iv) an evaluation framework for calibration, robustness, and bias on decentralized social media data. All work follows privacy and ethical guidelines, with no attempts to access private accounts or infer sensitive attributes beyond the scope of the data.

## 7 Conclusions and Future Work

We built a reproducible pipeline for studying sentiment/emotion signals and hashtag structure on Bluesky under decentralization constraints. The completed work establishes (i) descriptive differences across feed samples (language mix, engagement sparsity, reply structure, temporal activity), (ii) POLITISKY24 stance and hashtag summaries that expose politically driven tagging patterns, and (iii) a first modeling layer combining transformer inference with TF-IDF baselines under weak supervision. Together, these outputs provide a concrete starting point for evaluating transfer to Bluesky and for modeling tag groupings via graphs (RQ1-RQ2).

The main limitation of the current stage is evaluation: parts of the analysis rely on model-inferred labels, and we have not yet measured accuracy and calibration against gold sentiment labels on Bluesky. In addition, bias and fairness analysis (RQ3) requires either carefully defined group

proxies or controlled counterfactual probes, which are not yet implemented.

## 7.1 Future work

To complete RQ1-RQ3, the next steps are:

- **Gold-label evaluation for RQ1:** create or reuse a small human-annotated Bluesky subset to measure accuracy (macro-F1) and calibration (e.g., ECE and reliability curves) for transformer and baseline models.
- **Robustness for RQ1:** stress-test performance on Bluesky-specific phenomena (hashtags, emojis, slang, code-switching, and short/no-context posts), including ablations with and without tags.
- **Stability for RQ2:** formalize stability of tag groupings across feeds/instances using graph-community stability measures and embedding-based clustering stability.
- **Bias and mitigation for RQ3:** implement demographic or identity probes (or counterfactual templates), quantify group disparities in predictions and calibration, and evaluate mitigation strategies (reweighting, debiasing, post-hoc calibration).
- **Optional downstream analysis:** integrate sentiment/tag signals into a simple feed-ranking simulation to study potential social effects.

## 8 Code, Reproducibility, and Checklist

### 8.1 Repository structure

The code is organized as follows:

- `code/eda/EDA_Bluesky_social.ipynb`: feed-level EDA across multiple Bluesky feed samples.
- `code/eda/EDA_POLITISKY24.ipynb`: EDA for POLITISKY24 (stance/metadata + hashtag summaries).
- `code/eda/Labels_POLITISKY24.ipynb`: tag-centric analysis using predicted emotion labels and co-occurrence graphs.
- `code/ml_models/Sentiment_Analysis_ML_Models.ipynb`: transformer sentiment inference and TF-IDF baseline training.
- `presentation/`: milestone slides.
- `reports/`: proposal and milestone report versions.

### 8.2 Reproducibility checklist

- **Data access:** experiments rely on the Bluesky Social Dataset and POLITISKY24 (when applicable), plus locally collected feed samples. Any file paths must be updated to match the local environment.
- **Preprocessing:** notebooks document text normalization steps, timestamp standardization, and hashtag extraction rules. Language filtering is applied where needed (English-only for some modeling stages).
- **Models used:** sentiment inference uses `distilbert-base-uncased-finetuned-sst-2-english`; emotion inference uses `j-hartmann/emotion-english-distilroberta-base`; baselines include TF-IDF + Logistic Regression / Multinomial Naive Bayes / LinearSVC.
- **Training/evaluation splits:** classical baselines use stratified train/test splits and cross-validation as implemented in the ML notebook. Reported scores should always indicate whether labels are human gold or pseudo-labeled.
- **Randomness control:** for any train/test splits and model training, random seeds should be fixed (e.g., `random_state`) and reported in the notebook output or text.
- **Compute environment:** code is intended to run in Python with standard scientific libraries and HuggingFace Transformers. GPU is optional but speeds up transformer inference.
- **Outputs:** notebooks produce EDA tables/plots, extracted hashtag statistics, tag graphs, predicted label files (where saved), and baseline evaluation summaries. These outputs are expected to be reproducible given identical inputs.
- **Ethics and privacy:** analysis is restricted to public data and does not attempt to access private accounts. Any shared artifacts should avoid including personal identifiers beyond what is required for scientific reporting.

### References

- W. Zhang, Y. Deng, B. Liu, S. Pan, and L. Bing. 2024. Sentiment Analysis in the Era of Large Language Models: A Reality Check. *Findings of the As-*

- sociation for Computational Linguistics: NAACL. <https://doi.org/10.48550/arXiv.2305.15005>
- L. He, S. Omranian, S. McRoy, and K. Zheng. 2024. Using Large Language Models for Sentiment Analysis of Health-Related Social Media Data: Empirical Evaluation and Practical Tips. *medRxiv* preprint. <https://doi.org/10.1101/2024.03.19.24304544>
- M. Nasution et al. 2023. Benchmarking Open-Source Large Language Models for Sentiment and Emotion Classification in Indonesian Tweets. University of Islam Riau repository. <https://doi.org/10.1109/ACCESS.2025.3574629>
- X. Fu et al. 2023. Efficacy of ChatGPT in Cantonese Sentiment Analysis: Comparative Study *PubMed*-indexed journal. <https://doi.org/10.2196/51069>
- M. Kleppmann et al. 2024. Bluesky and the AT Protocol: Usable Decentralized Social Media. *ACM* <https://doi.org/10.1145/3694809.3700740>
- E. Sahneh, G. Nogara, M. DeVerna, N. Liu, L. Luceri, F. Menczer, F. Pierri, and S. Giordano. 2025. The Dawn of Decentralized Social Media: An Exploration of Bluesky's Public Opening. ISBN: 978-3-031-78540-5. pp.422-437 [https://doi.org/10.1007/978-3-031-78541-2\\_26](https://doi.org/10.1007/978-3-031-78541-2_26).
- T. Huang. 2024. Decentralized social networks and the future of free speech online. *Computer Law & Security Review*, 55:106059. <https://doi.org/10.1016/j.clsr.2024.106059>.
- Y. Djenouri, A. Belhadi, and J. C. W. Lin. 2019. Deep learning based hashtag recommendation system for multimedia data. *Information Processing & Management*. <https://doi.org/10.1016/j.ins.2022.07.132>
- H.-T. Liou et al. 2020. Dynamic Graph Transformer for Implicit Tag Recognition. In *Proceedings of ACL*. <https://doi.org/10.18653/v1/2021.eacl-main.122>
- W. Feng et al. 2015. STREAMCUBE: Hierarchical spatio-temporal hashtag clustering for event exploration over the Twitter stream. *Proceedings - International Conference on Data Engineering*, 2015, 1561-1572. <https://doi.org/10.1109/ICDE.2015.7113425>.
- X. Jin et al. 2024. MM-Soc: A Comprehensive Benchmark for Multimodal LLMs on Social Media. In *Proceedings of ACL*. <https://doi.org/10.48550/arXiv.2402.14154>
- Q. Pan and Z. Meng. 2024. Hybrid Uncertainty Calibration for Multi-modal Sentiment Analysis. *Electronics*. <https://doi.org/10.3390/electronics13030662>.
- T. Xiao et al. 2022. Uncertainty Quantification and Calibration for Pre-Trained Language Models. In *Findings of ACL*. <https://doi.org/10.48550/arXiv.2210.04714>
- M. I. Radaideh, O. H. Kwon, and M. I. Radaideh. 2025. Fairness and Social Bias Quantification in Large Language Models for Sentiment Analysis. *Knowledge-Based Systems*, 319:113569. <https://doi.org/10.1016/j.knosys.2025.113569>.
- J. P. Venugopal, A. A. Subramanian, G. Sundaram, M. Rivera, and P. Wheeler. 2023. A Comprehensive Approach to Bias Mitigation for Sentiment Analysis of Social Media Data. *Applied Sciences*, 14(23):11471. <https://doi.org/10.3390/app142311471>.
- S. Vallejo Vera and H. Driggers. 2025. LLMs as Annotators: The Effect of Party Cues on Labelling Decisions by Large Language Models. *Humanities and Social Sciences Communications*, 12:1530. <https://doi.org/10.1057/s41599-025-05834-4>.
- R. Corizzo and F. S. Hafner. 2024. Mitigating Social Bias in Sentiment Classification via Ethnicity-Aware Algorithmic Design. *Social Network Analysis and Mining*, 14:208. <https://doi.org/10.1007/s13278-024-01369-9>.
- A. Kadriu et al. 2022. Human-annotated dataset for social media sentiment analysis for Albanian language. *Diva Portal* technical report. <https://doi.org/10.1016/j.dib.2022.108436>
- P. Rostami, V. Rahimzadeh, A. Adibi, and A. Shakeri. 2025. POLITISKY24: U.S. Political Bluesky Dataset with User Stance Labels [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.15616911>.
- A. Failla and G. Rossetti. 2025. Bluesky Social Dataset [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.14669616>.
- F. N. Silva, K.-C. Yang, W. Zhao, and B. Tran Truong. 2024. Data for: Exploring Emerging Social Media: Acquiring, Processing, and Visualizing Data with Python and OSoMe Web Tools [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.12748042>.