Unlocking Transparency in Platforms' Content Moderation Activities

DIGITAL SERVICES ACT

Introducing *dsa_tdb*, a Python Package for Analyzing the Digital Services Act Transparency Database

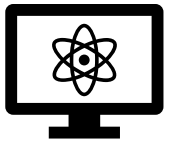
Overview of the DSA

The Digital Services Act aims at creating a safer digital space:

Transparent Terms and Conditions

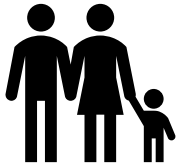
Consumer protection and personal rights

New transparency and data access provisions



Explain, risk assess & test algorithms (Rec. Sys.)

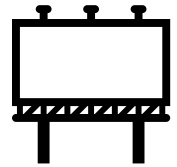
Content moderation policies



Protection of minors (no targeting)

Notice and action to report illegal content

Seller information



Advertisement details

Transparency features of the DSA

Transparency Reports

Bi(annual) statistics on content moderation, incl. accuracy, speed & human resources

Transparency Database

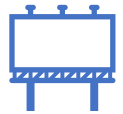
near real time content moderation seismometer

Terms & Conditions

OPEN TERMS ARCHIVE

Clear and Transparent Language of the T&C

Ad Library



Repository of the ads hosted by the platform

Risk Assessments

Analysis of algorithmic risk factors

Independent Audits

Test of algorithmic systems

Data Access

Study of systemic risks

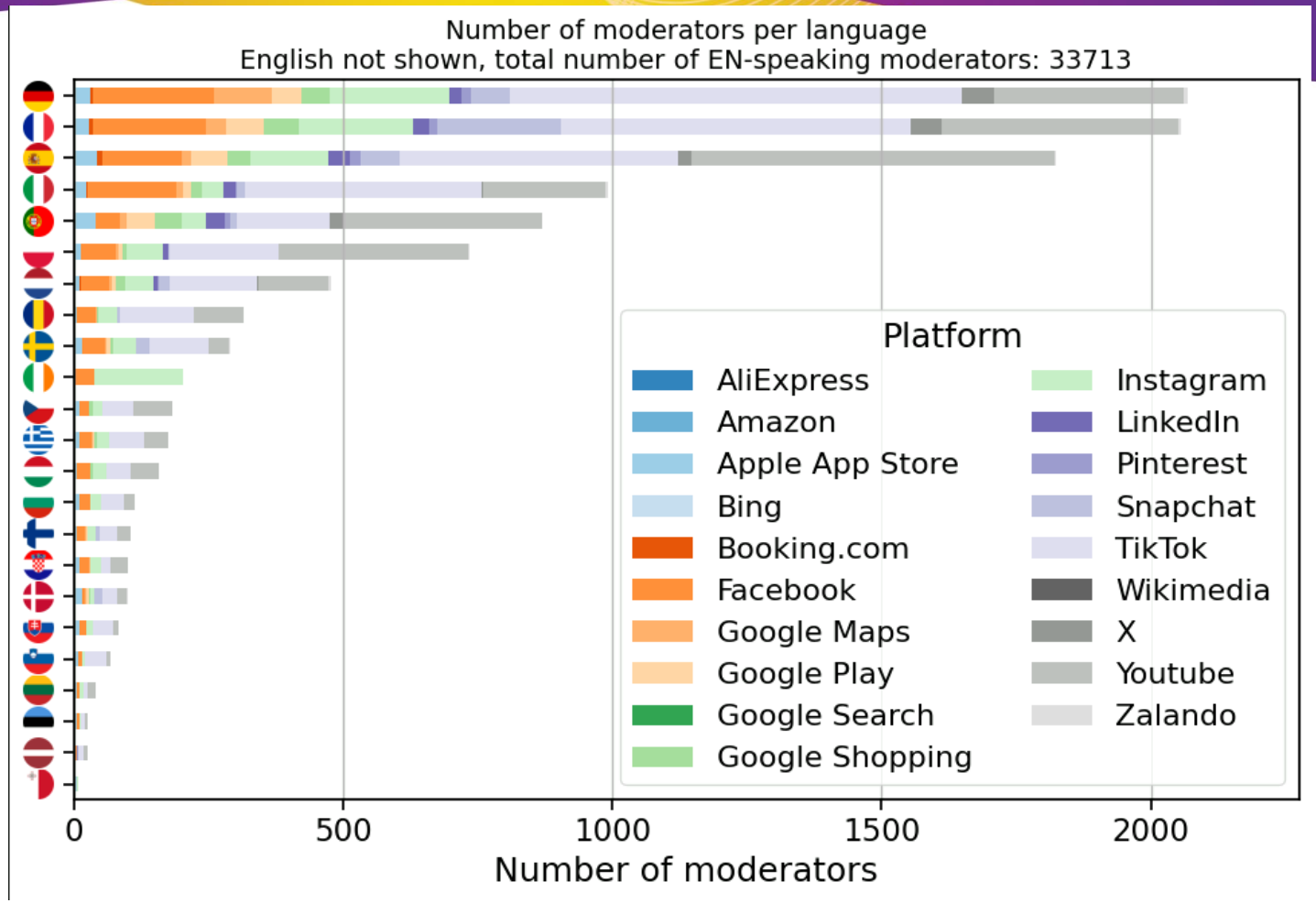
Whistleblower tool

Employee and stakeholders can anonymously report bad practices and infringments

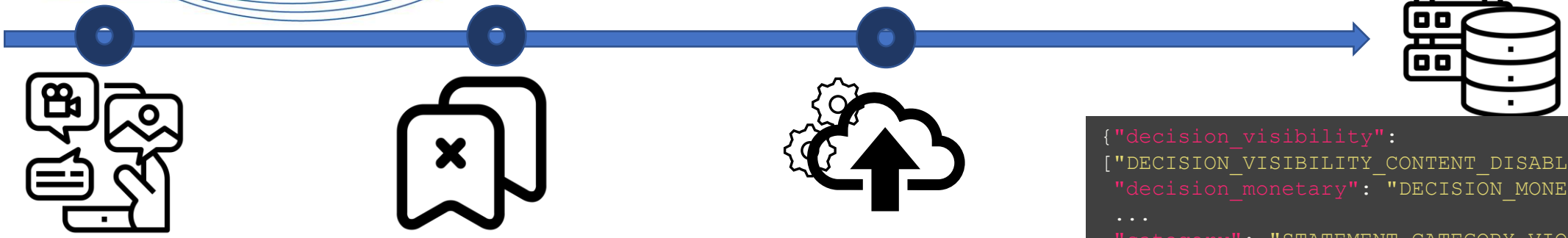
Transparency Reports

- Monthly Active Users (MAU)
- Content moderation Human resources
- Accuracy of content moderation and appeal

So far, 3 rounds of VLOPs Transparency Reports (Nov 23, May and Nov 24)



Transparency DB: data lifecycle

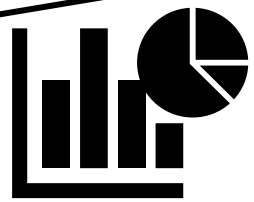


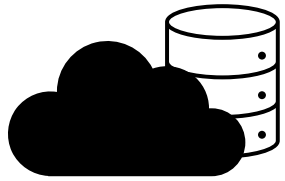
User creates content
`content_date`

Platform moderates content
`application_date`

Platform sends SoR to user & (anon) to TDB
`created_at`

```
{"decision_visibility":
["DECISION_VISIBILITY_CONTENT_DISABLED"],
 "decision_monetary": "DECISION_MONETARY_TERMINATION",
 ...
 "category": "STATEMENT_CATEGORY_VIOLENCE,
 "automated_detection": "No",
 "automated_decision": "AUTOMATED_DECISION_PARTIALLY"}
```





From website search

- Near Real time (24h delay)
- Limited to 10’000 (1’000 csv)
- **Only last 6 months**

Online dashboard

- Near real time (24h delay)
- Aggregate view of the data
- No export / no customisaton

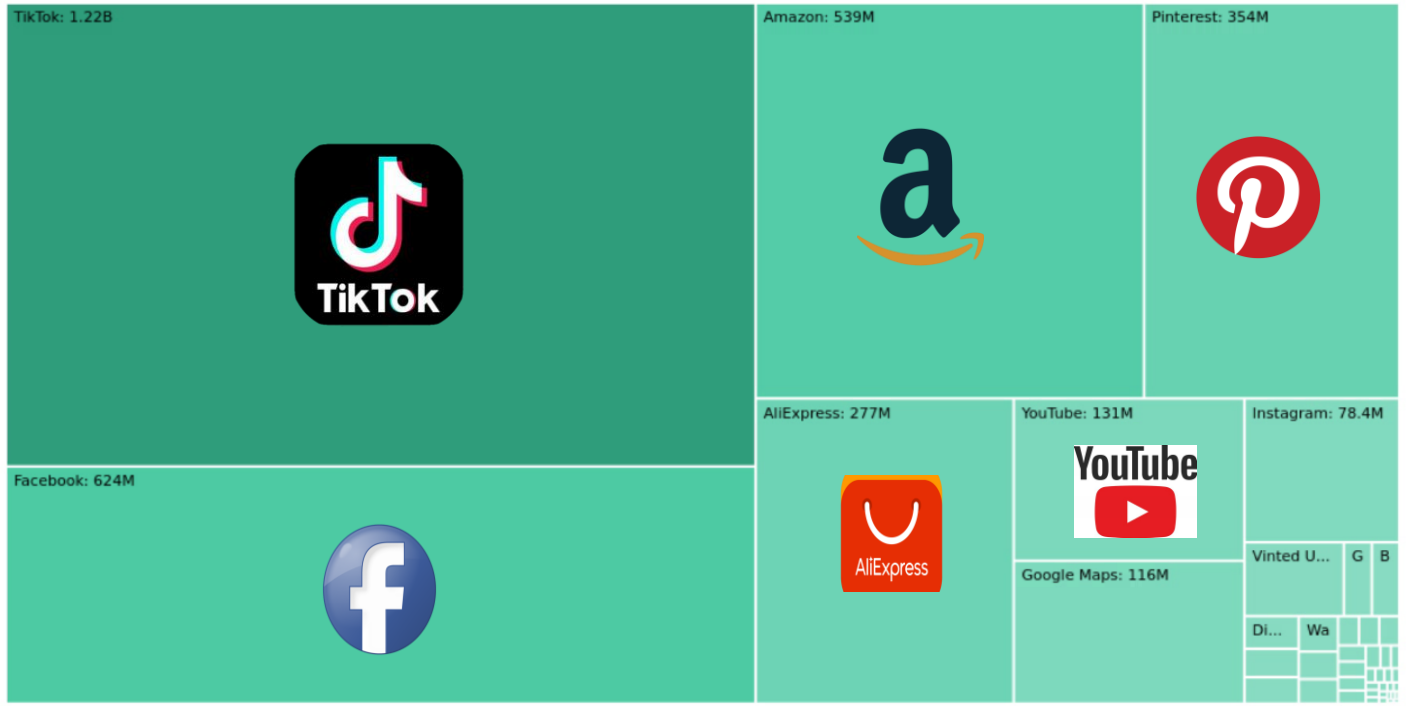
Daily dumps

- Near real time (24h delay)
- Contain CSV with full SOR details
- Large files + pre-/post-processing

Transparency DB: a massive dataset

The Transparency database is **large** (and *dsa-tdb* cannot do miracles).

- 1TB of **full daily dump** files
- 1.8GB of the aggregated dataset.
- ~5-10GB per day of caching when analysing.



dsa-tdb: how to use it

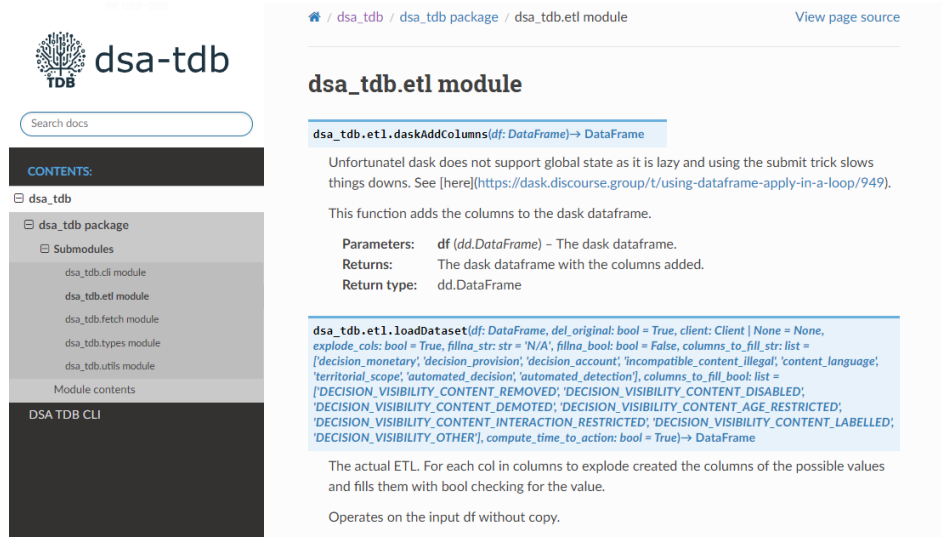
Three venues:

- ***pip install --index-url https://code.europa.eu/api/v4/projects/943/packages/pypi/simple dsa_tdb***

- ***Docker/Podman container***

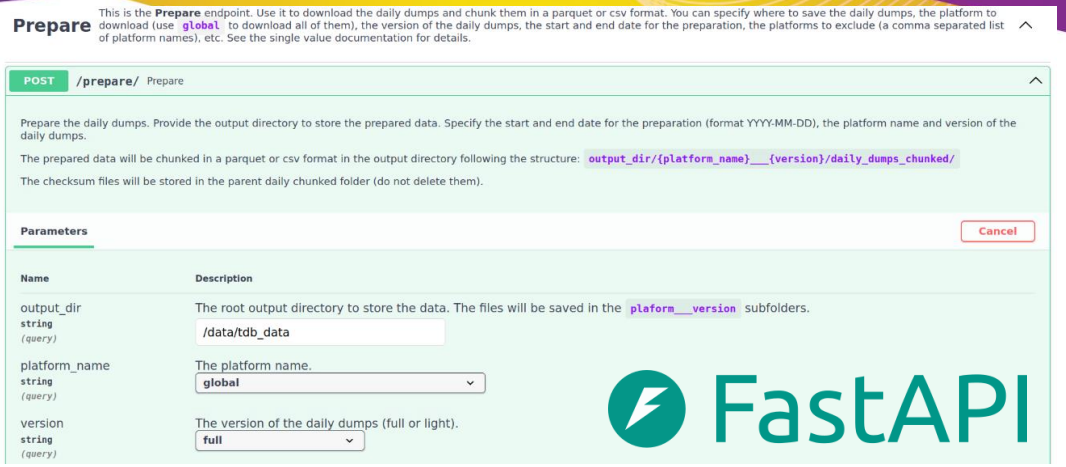
- ***Superset dashboards***

+ Online documentation:



dsa-tdb: API and cli/package

- API interface (download, filter, aggregate)



- CLI interface (download, filter, aggregate)

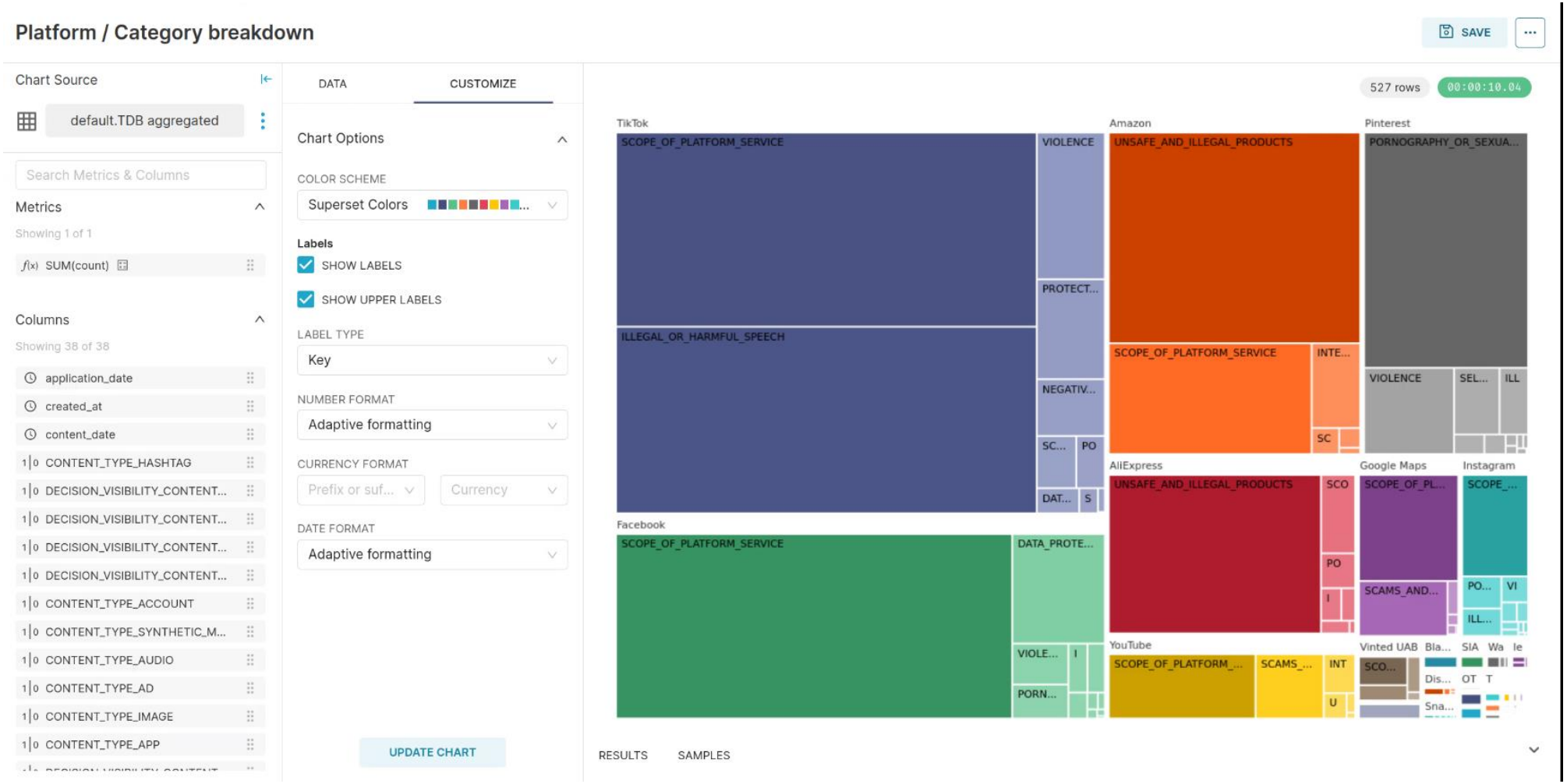
```
$ dsa-tdb-cli preprocess -p global
$ dsa-tdb-cli filter -c config.yaml
$ dsa-tdb-cli aggregate -c config.yaml
```

- + Interactive

Stay around and join the workshop for a demo and additional details!

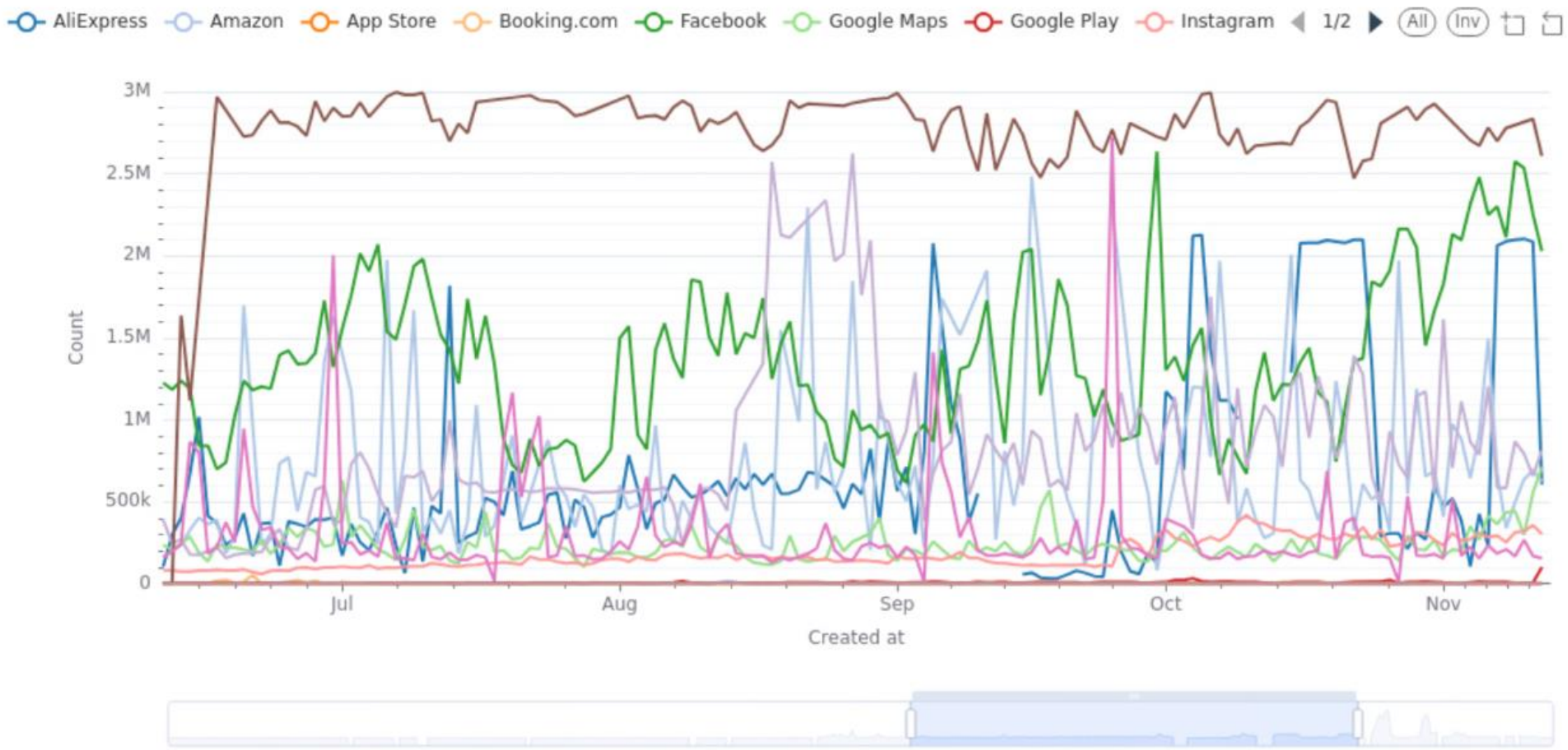
dsa-tdb: dashboard

Breakdown of platform and category



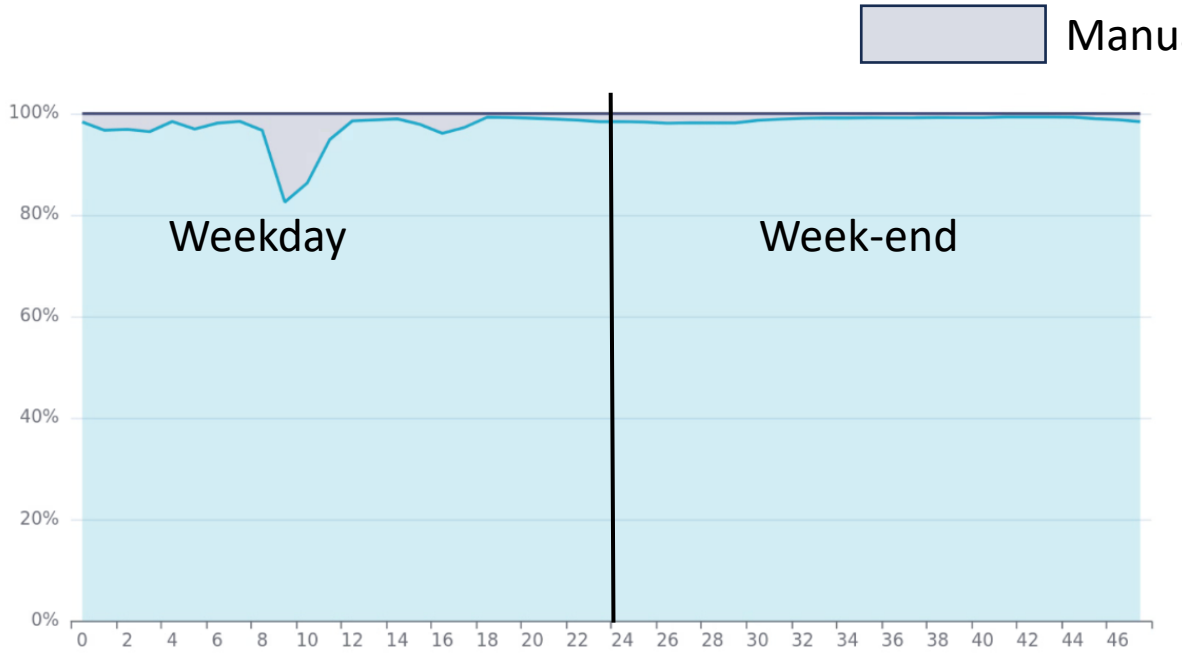
dsa-tdb: dashboard

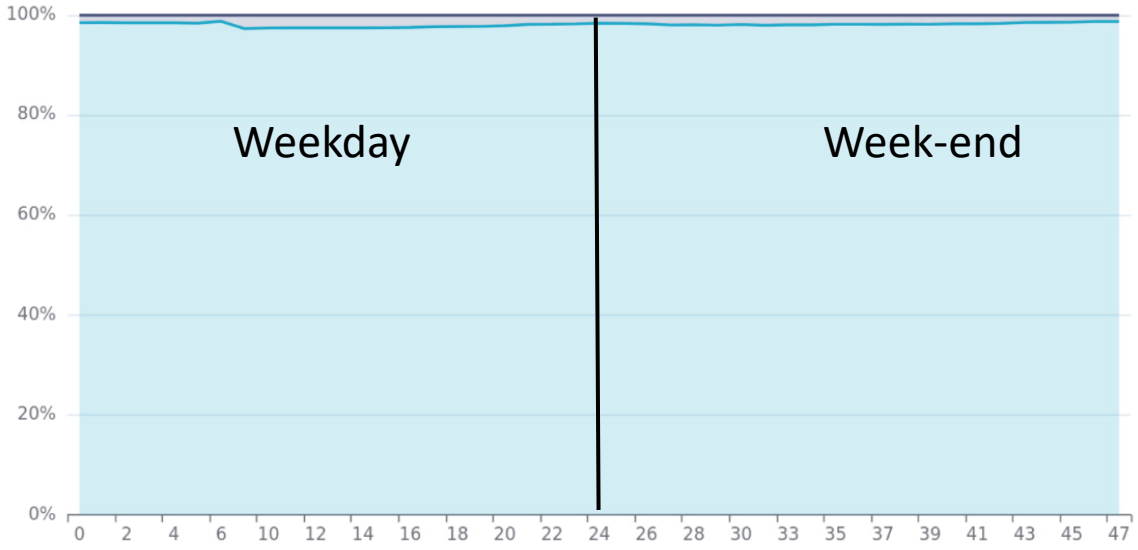
Daily submission volume by platform



dsa-tdb: dashboard

Automated or manual content detection





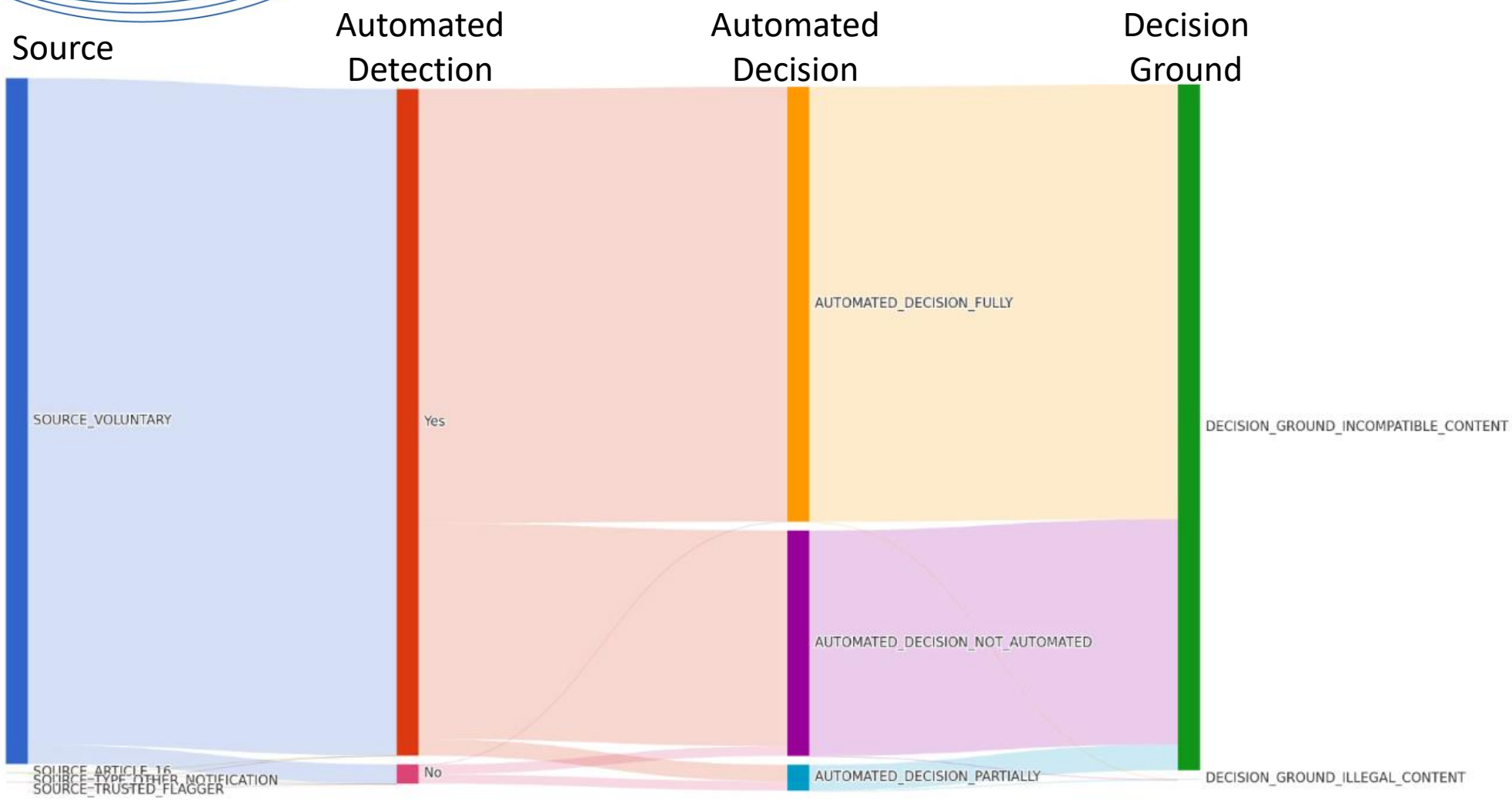
TikTok

=

+

dsa-tdb: dashboard

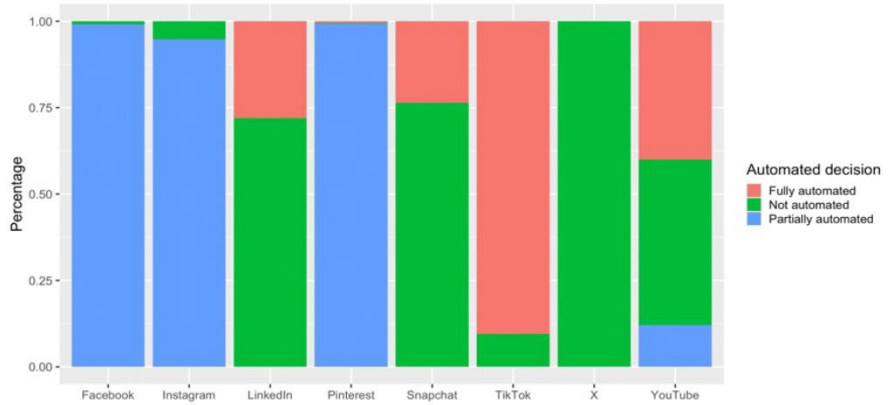
Flow of: Source

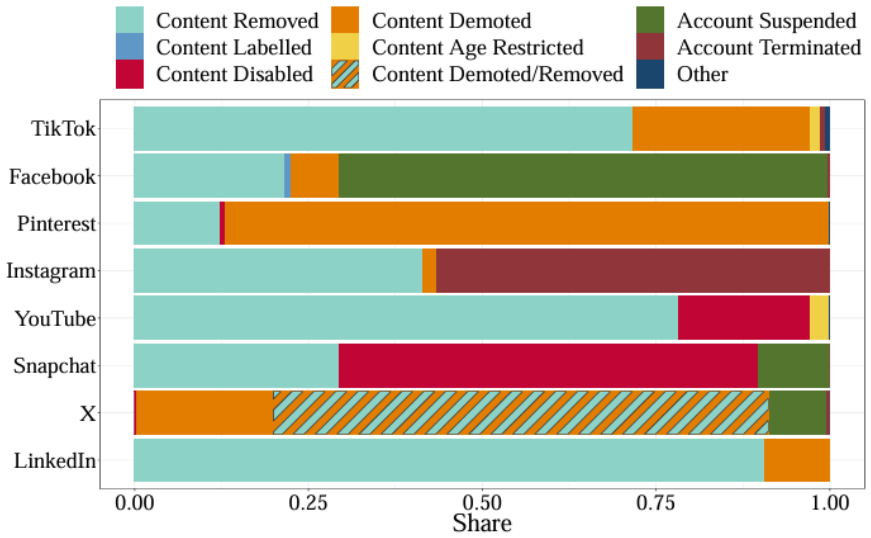


Community driven & research work on the Transparency DB

- Adaptation of the schema of the Database to the Reports in July 2025
  - Output of an open consultation between CSOs, researchers and companies
  - Optimizes reporting and aligns the transparency reporting provisions
  - Adds a product identifier (EAN-13) to track the spread of illegal products online
- Flourishing research community
  - Kaushal, R., et al., Automated Transparency: A Legal and Empirical Analysis of the Digital Services Act Transparency Database. Preprint at https://doi.org/10.48550/arXiv.2404.02894 (2024).
  - Drolsbach, C. & Pröllochs, N., Content Moderation on Social Media in the EU: Insights From the DSA Transparency Database. Preprint at https://doi.org/10.48550/arXiv.2312.04431 (2023).
  - Dergacheva, D., et al. One Day in Content Moderation: Analyzing 24 h of Social Media Platforms' Content Decisions through the DSA Transparency Database. (2023) doi:10.26092/elib/2707.
  - Platforms overwhelmingly use automated content moderation, first DSA transparency reports show – Lab Platform Governance, Media and Technology (PGMT). https://platform-governance.org/2023/platforms-overwhelmingly-use-automated-content-moderation-first-dsa-transparency-reports-show/ (2023).
  - Trujillo, A., Fagni, T. & Cresci, S. The DSA Transparency Database: Auditing Self-reported Moderation Actions by Social Media. Preprint at https://doi.org/10.48550/arXiv.2312.10269 (2023).
  - Miller, G. Tracking the First Digital Services Act Transparency Reports | TechPolicy.Press. *Tech Policy Press* https://techpolicy.press/tracking-the-first-digital-services-act-transparency-reports (2023).

Dergacheva et al.: One Day in Content Moderation





Drolsbach, C. & Pröllochs, N.: Content Moderation on Social Media in the EU

DSA Transparency DB

Backend github.com/digital-services-act/transparency-database

Website transparency.dsa.ec.europa.eu

Package code.europa.eu/dsa/transparency-database/dsa-tdb

Open source -> Try it, open issues and pull-requests are welcome!

dsa-tdb

Stay around and join the workshop!

Thank You